Improving Bayesian VAR density forecasts through autoregressive Wishart Stochastic Volatility*

Paul Karapanagiotidis†

Thesis draft
June 22, 2014

Abstract

This paper considers a vector autoregressive model (VAR) model with stochastic volatility which appeals to the Inverse Wishart distribution. Dramatic changes in macroeconomic time series volatility pose a challenge to contemporary VAR forecasting models. Traditionally, the conditional volatility of such models had been assumed constant over time or allowed for breaks across long time periods. More recent work, however, has improved forecasts by allowing the conditional volatility to be completely time variant by specifying the VAR innovation variance as a distinct discrete time process. For example, Clark (2011) specifies the elements of the covariance matrix process of the VAR innovations as linear functions of independent nonstationary processes. Unfortunately, there is no empirical reason to believe that the VAR innovation volatility processes of macroeconomic growth series are nonstationary, nor that the volatility dynamics of each series are structured in this way. This suggests that a more robust specification on the volatility process—one that both easily captures volatility spill-over across time series and exhibits stationary behaviour—should improve density forecasts, especially over the long-run forecasting horizon. In this respect, we employ a latent Inverse Wishart autoregressive stochastic volatility specification on the conditional variance equation of a Bayesian VAR, with U.S. macroeconomic time series data, in evaluating Bayesian forecast efficiency against a competing specification by Clark (2011).

*I'd like to thank John Maheu, Christian Gourieroux, Angelo Melino, Martin Burda, and Carolyn Pitchik for their helpful comments and suggestions.

†University of Toronto, Department of Economics, p.karapanagiotidis@utoronto.ca

Keywords: Inverse Wishart distribution, stochastic volatility, predictive likelihoods, MCMC, macroeconomic time series, density forecasts, vector autoregression, steady state priors, Bayesian econometrics
JEL: C11, C32, C53, E17

1 Introduction

Forecasts of macroeconomic time series have become a ubiquitous component of any policymaker's toolkit. As such, central banks like the Federal Reserve typically publish density forecasts for inflation, output, interest rates, or other major indicators. This information helps both industry and consumers make decisions consistent with economic fundamentals. However, forecasts themselves are not infallible. In fact, while major advances have been made in the area of statistical forecasting, there remains much room for improvement.

This paper resolves some of the relevant issues by proposing a key change in the volatility process of Vector Autoregressions (VAR) popular among macroeconomists. Instead of assuming that the time varying VAR innovation covariance structure is driven by independent nonstationary processes, we employ a stationary multivariate Inverse Wishart process where the scale matrix is a function of past covariance matrices. Furthermore, we employ four major U.S. macroeconomic data series, that are, the rate of GDP growth, the inflation rate, the interest rate, and the unemployment rate, respectively.[1] A Bayesian approach, employing Markov Chain Monte Carlo methods (MCMC), is then taken in both estimation and in comparing forecasts between the benchmark model [Clark (2011)] and our competing Inverse Wishart autoregressive

[1]Note that the data is taken from the RTDSM database – the same dataset, in fact, as Clark (2011), our benchmark comparison model (with the exception of the interest rate—see Section 2 on the data set, below). Moreover, all data is at the aggregate U.S. level. Finally the interest rate employed in our paper is the 3-month Federal Treasury Bill rate.

volatility specification [Philipov and Glickman (2006)]. Results suggest that incorporating the more sophisticated Inverse Wishart autoregressive volatility process can improve density forecasts in both the short and long-run, with larger improvements as the horizon increases, despite a small data set and increased parameterization of the model. With this in mind, the following discussion aims to provide a broader context surrounding the relevant forecasting issues precipitating this proposed modification to the typical VAR process volatility specification.

1.1 Background

A fundamental issue facing the production of good forecasts has been that of how to deal with the changing moments of the conditional forecast distributions. For example, dramatic changes in U.S. economic volatility have posed a modeling challenge to contemporary forecasters, specifically among macroeconomists where Gaussian VAR models are popular. An analysis of major U.S. economic indicators, such as output growth over the past 100 years, illustrates that the economy goes through periods of changing volatility. For example, "The Great Moderation," which began in the 1980's, represented a period of unusually low volatility vis-a-vis both a lengthy prior period of erratic volatility and the more recent instability we've experienced since 2007. In this respect, both Sims (2001) and Stock (2001), in separate discussions of Cogley and Sargent's (2001) paper, criticized the assumption of homoskedastic VAR variances, pointing to evidence analyzed by Bernanke and Mihov (1998a,1998b), [2] Kim and Nelson (1999), or McConnell and Perez Quiros (2000). [3] Clark (2011) also finds significant changes in conditional volatility across time when estimating the latent stochastic volatilities of a

[2] In the case of monetary policy shocks between 1979 and 1982.
[3] With respect to the growing stability of output around 1985.

Bayesian VAR model.

It should not come as a surprise then, that while, traditionally, the volatility of forecasting models was assumed constant over time (primarily for the sake of simplicity), it can be shown that this assumption leads to poor conditional forecasts. For example, Jore, Mitchell, and Vahey (2010) employ a model averaging approach to U.S. data, with both equal weights and recursively adapted weights, based on log predictive density scores across a range of different specifications. Their results show strong support for a recursive weighting scheme across specifications. More interestingly, however, they find that during periods of changing macroeconomic volatility, for example when the U.S. economy transitioned into "The Great Moderation", the weighting scheme tends to place more weight on specifications which dynamically account for structural breaks in volatility. Moreover, they find evidence of poor forecasting given a simple assumption of fixed volatility or equal weights across model specifications. However, it worth noting that the specifications which do respond to structural change within Jore et al.'s (2010) framework are limited in that they are restricted to a finite set of possible volatility states and breaks.

Consequently, it is important to account for changing volatility in any forecasting specification. Furthermore, if such changes in volatility occurred relatively infrequently and could be extracted from the data with reasonable statistical significance, then employing a regime switching specification such as in Jore et al. (2010) might prove sufficient in drawing good forecasts. However, the truth is that, given the complexity of the economy, changes in volatility probably occur much more frequently and take on many more values than can be effectively captured by a finite state-space model. For this reason forecasters have adopted a continuous state-space framework for estimating the conditional volatility of VAR models as opposed to the finite state regime switch-

ing type model applied to volatility, as popularized by Hamilton (1989), and employed by Jore et al (2010). Moreover, the use of the so-called continuous state "Stochastic Volatility" model has also grown in popularity given its usefulness in modeling a latent volatility process based on a filtration that includes more than just lagged VAR series shocks, as for example in the case of a GARCH model or a volatility-in-mean model.[4]

Both Cogley and Sargent (2005), and Primiceri (2005), allow for time variation in the conditional covariance matrix across VAR series shocks according to a Stochastic Volatility law of motion, where the conditional volatility can take on any value in a continuous positive real set (and covariances can be any real number). Moreover, they also allow for time variation in the VAR parameters themselves, through another Stochastic Volatility law of motion on their state across time. Clark's (2011) model, which represents our benchmark, follows the same structure of the previous two studies, albeit without the time varying VAR parameters, which are dropped in favour of tight Bayesian steady state priors on the deterministic trend parameters (which define the unconditional mean of the VAR process) and a rolling sample window which re-estimates the parameters across time.

Villani (2009) showed that imposing Bayesian steady state prior distributions allow us to incorporate prior beliefs about macroeconomic variable steady states into our model. Furthermore, our belief is that employing this information probably reduces the need for time varying VAR parameters since much of the time variation in the autoregressive parameters (which is not due to a lack of time variation in the shock covariance, as was the case with Cogley and Sargent (2001)), may in fact be due to a lack of a well defined trend (see Cogley and Sargent (2005), where they model their VAR intercepts[5]

[4]See also Sartore and Billio (2005) for a useful survey of Stochastic Volatility.

[5]Noting of course that given their formulation, the VAR long-run mean μ_t is both time varying, stochastic, and a function of the VAR intercepts, α_t, as $\mu_t = (\mathrm{I} - \Phi_t)^{-1}\alpha_t$.

as stochastic random walks). Moreover, given the quarterly nature of most macroeconomic time series, small sample sizes are usually the norm. In this situation a tight prior also plays the role of constraining VAR parameters to aid in the identification of trends that might not otherwise be readily apparent.[6] In this respect, Villani (2009) demonstrates that informative priors for trends can greatly improve point forecasts, especially over the longer term horizon where correct specification of the trend of the series is important—see Clements and Hendry (1998). All of this of course assumes that our prior beliefs on the nature of the time series trends are the right ones. In fact, whether or not the trends in macroeconomic data are better modeled as stochastic (i.e. unit roots with drift) or deterministic is still an open question of debate.

However, most of these studies adopt certain features which could still be improved upon. For example, many of these studies construct the VAR innovation covariance dynamics as driven by independent nonstationary processes. Often, some form of fixed relationship is imposed between the elements of the VAR innovation covariance matrix and the independent processes driving them. This is done in order to reduce the parameterization of the model, but it may limit the richness of the covariance matrix dynamics.

The empirical rationale for this choice of specification is not entirely clear since an analysis of many macroeconomic time series suggest stationary volatility dynamics. Moreover, without explicitly parameterizing time varying covariances matrices, it is extremely difficult to interpret any volatility spill-over effects, as have been shown to be prevalent among financial time series from U.S. markets (see for example, Diebold and Yilmaz, (2008)). Furthermore, studies such as Cogley and Sargent (2005) seem to provide little explicit justification for the choice of a nonstationary process driving VAR

[6] In Clark (2011) for example, his rolling sample window is only of size $T = 80$.

innovation volatility, other than a brief comment that "the random walk specification is designed for permanent shifts in the innovation variance, such as those emphasized in the literature on the growing stability of the U.S. economy." Ultimately, if the assumption of nonstationarity is misspecified and there does in fact exist volatility spill-over across macroeconomic time series, then these existing specifications leave something to be desired.

Given this, the multivariate volatility process can be constructed to directly model the time varying covariance matrices without simply extending, in an ad hoc way, the traditional univariate specification to the multivariate case. Moreover, any autoregressive persistence in volatility can be captured and a finite unconditional mean can be specified. Philipov and Glickman (2006) [see also Chib, Omori, and Asai (2009)] apply such an autoregressive Inverse Wishart process to analyze the conditional volatility of financial data and find that it improves volatility forecasts over simpler formulations, where a number of Bayesian and frequentist measures are applied to compare forecast accuracy given a variety of competing specifications. It is worth noting, however, that there exist problems with the Philipov and Glickman (2006) implementation of the Inverse Wishart autoregressive volatility process as it stands—see Rinnergschwentner et al (2011) for more details and quite a few corrections. Therefore, in light of Phillipov and Glickman's idea of Inverse Wishart covariance modeling, we propose a modified version of their process which we feel serves the purpose of multivariate forecasting better than Clark (2011)—see Section 3 below for more details, including differences between our model and that of Philipov and Glickman (2006).

The rest of the paper is organized as follows. Section 2 discusses the data and the methodology used to adjust the data for trends. Section 3 discusses both the benchmark model based on Clark (2011) and the proposed Inverse Wishart process modification

based on Philipov and Glickman (2006). Section 4 then details the trend specification and the conjugate priors we impose within the Bayesian framework. Section 5 then discussion estimation of the model parameters by Gibbs sampler method. Section 6 discusses the method whereby we generate forecast densities for both the VAR levels and covariance matrices across various horizons. Section 7 details the results of both the estimation process and the forecast comparisons based on Bayesian analysis of the predictive likelihoods. Finally, Section 8 summarizes and concludes.

2 Data

We consider four macroeconomic time series generated from aggregate U.S. data, that are

1. the real output growth,
2. a measure of the inflation rate,
3. the unemployment rate,
4. and an interest rate.

The data source is the same as in Clark (2011): the so-called "real-time" [7] data from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomics (or "RTDSM"). The total sample size is quite small: only $T = 252$ data points extending from the 2^{nd} quarter of 1948 (hereon denoted as 1948:Q2) until the 1^{st} quarter of 2011. Output from the RTDSM database is quarterly real data and measured as either

[7] That is data that is regenerated annually to conform to new changes in the way we measure macroeconomic indicators, or to take into account flaws in some previous set, observed ex-post. Each new issue is deemed a "vintage."

Gross Domestic Product (GDP) or Gross National Product (GNP) depending on the data vintage.[8] Inflation from the RTDSM is also measured quarterly and as either a GDP or GNP deflator or a price index, depending on the vintage. We measure growth and inflation rates as annualized log changes.[9] The unemployment rate, however, is available on a monthly basis so we simply average across each quarter in matching the quarterly nature of output and inflation. Moreover, it should be noted that the unemployment rate tends to differ much less dramatically across vintages. Finally, while Clark (2011) employs the federal funds interest rate series, Primiceri (2005) recommends the nominal annualized yield on 3-month Federal Treasury Bills, since this series goes back much further. We therefore adopt the latter series, and again, average across quarters since the data is monthly.[10] Finally, output, inflation, and the unemployment rate are already seasonally adjusted by their source providers.

Clark and McCracken (2008,2010) also provide evidence that point forecasts of GDP growth, inflation, and interest rates are improved by specifying the latter two series as deviations from some form of deterministic trend simulating inflation expectations. Given this result Clark (2011) adopts the Blue Chip Consensus forecast produced from survey data and published by Aspen Publishers Ltd., as a form of long-term inflations expectations. Unfortunately, as Clark mentions in his online appendix, the data for this Blue Chip forecast of inflation expectations only extends back to the fourth quarter of

[8]The RTDSM generates entirely new time series each quarter (deemed "vintages") based on updated chain weighting techniques or other improvements. Thus newer vintages represent larger samples than older ones which were generated at previous dates.

[9]Since log differences are already continuously compounded, we simply multiply each quarterly value by 4.

[10]The 3-month Federal Treasury bill rate series employed is a combination of two very similar series joined together at June 2000, since the first vintage was discontinued. "H15/discontinued/H1.RIFSGFPIM03_N.M" is the unique ID for the discontinued series and "H15/H15/RIFLGFCM03_N.M" is the newer series. Both series are available at the Federal Reserve website: `http://www.federalreserve.gov/releases/h15/data.htm`.

1979 (i.e. 1979:Q4). Therefore, Clark appends an exponentially smoothed trend from his inflation series to be beginning of the Blue Chip series in extending it back to 1964. Clark mentions that despite his attempts at keeping the data "as real time as possible" by employing every quarterly vintage of inflation data, in the end a trend based on his most recent vintage (2008:Q4) deviates little from the others. Moreover, as Clark notes, Kozicki and Tinsley (2001a,2001b) and Clark and McCracken (2008), both suggest that exponentially smoothed trends of the inflation rate match up reasonably well with survey-based measures of long-run expectations in the data since the early 1980's. Given both of these facts, we will simply employ an exponentially smoothed trend of the inflation rate through the most recent vintage currently available (2011:Q4) in generating a long-term inflation expectations series, skipping the Blue Chip survey data entirely and ignoring the previous vintages of inflation data.[11]

Finally, the unemployment rate series is also detrended by an exponential smoother (in the same way the inflation rate was detrended in order to generate the long-run inflation expectations [see footnote 11]).

Therefore, to summarize:

1. GDP growth is not detrended (but will be centered on a long-run constant mean of 3.0% through the prior distribution).

2. The inflation rate is detrended by its exponentially smoothed trend (with a smoothing parameter of $\alpha = 0.05$).

3. The interest rate (3-month Treasury bill) is detrended around the same trend as in-

[11] The exponential smoother employed is as follows: $y_t^* = y_{t-1}^* + \alpha(y_t - y_{t-1}^*)$, where y_t is the actual data series and y_t^* is the exponentially smoothed trend. α is a parameter which can be adjusted depending on how "tight" we want the trend to follow the data series. For the inflation rate trend used as long-term inflation expectations, Clark suggests a value of $\alpha = 0.05$.

flation (which is supposed to simulate long-term inflation expectations), although we force a long-run constant mean of 2.5% above trend through the prior on the unconditional mean.

4. The unemployment rate is detrended by its exponentially smoothed values lagged one period, with a smoothing parameter of $\alpha = 0.02$.

See the model and estimation Sections 3 and 5 respectively, for more details as to how these trends are implemented into the model.

3 Model specifications

The benchmark model is the Bayesian VAR, steady-state prior, Stochastic Volatility specification (BVAR-SSP-SV) as outlined in Clark (2011). This model employs a Bayesian $VAR(J)$ formulation for the detrended series, where the covariance matrix of the VAR innovations is driven by linear functions of separate univariate, independent, geometric Brownian motions.

3.1 Benchmark model

We refer to this benchmark model from Clark (2011) as the $Clark$ specification:

$$\mathbf{v_t} = \mathbf{\Pi}\left(L\right)\left(\mathbf{y_t} - \mathbf{\Psi d_t}\right), \tag{1a}$$

$$\text{where} \quad \mathbf{\Pi}\left(L\right) = \mathbf{I_p} - \sum_{j=1}^{\mathbf{J}} \mathbf{\Pi_j} L^j \quad \text{and} \;\; L \;\; \text{is the lag operator,} \tag{1b}$$

$$\mathbf{v_t} = \mathbf{B}^{-1}\mathbf{\Lambda}_t^{0.5}\varepsilon_{\mathbf{t}} \quad \text{where} \quad \varepsilon_{\mathbf{t}} \sim MVN_p\left(\mathbf{0}, \mathbf{I_p}\right), \tag{1c}$$

$$\mathbf{\Lambda_t} = diag(\lambda_{1,t}, \lambda_{2,t}, \ldots, \lambda_{p,t}), \tag{1d}$$

$$\ln\left(\lambda_{i,t}\right) = \ln\left(\lambda_{i,t-1}\right) + \xi_{i,t}, \quad \forall i = 1, \ldots, p, \tag{1e}$$

$$\text{and} \quad \xi_{i,t} \sim i.i.d. \;\; N(0, \varphi_i), \quad \forall i = 1, \ldots, p. \tag{1f}$$

The first equation, (1a), is a vector autoregressive model applied to the macroeconomic series, adjusted for trends. The trends have to be estimated by means of the $\mathbf{\Psi}$ matrix. More precisely, we introduce the state variables $\mathbf{d_t}$ as in Villani (2009). The state variables can be chosen in a number of ways.[12] The $\mathbf{y_t}$ admits a time varying unconditional mean, $\mu_{\mathbf{t}} = \mathbf{\Psi d_t}$, where $\mathbf{\Psi}$ is a $p \times q$ matrix and $\mathbf{d_t}$ is a $q \times 1$ vector of deterministic trends.

$\mathbf{\Pi}(L)$ is the matrix lag-polynomial and $\mathbf{v_t}$ denotes the innovations of the process $(\mathbf{y_t})$. The second equation, (1c), highlights the form of the stochastic volatility of the

[12]For example, if $\mathbf{d_t} = 1, \forall t$, i.e. takes on a single constant value for all time periods, then $\mathbf{\Psi}$ is a vector of regression constants, the values of which determine the time invariant long-run means of the autoregressive levels processes, $\mathbf{y}_t$. However, if for example, $\mathbf{d_t} = t$, then the elements of the vector $\mathbf{\Psi}$ represent the slope coefficients of a linear time-trend relationship shared by each of the $\mathbf{y_t}$ series. Furthermore, if $\mathbf{d_t} = [t, f(t)]$ for example, where $f(t)$ is perhaps some nonlinear function of t, then $\mathbf{\Psi}$ is now a matrix of "factor loadings," the elements of which reflect how the time varying long-run means of each process are expressed as a linear combination of both the linear and nonlinear time trend, simultaneously. This approach to modeling the unconditional levels first-order moment allows for greater flexibility. For example, we could incorporate a pre-exponentially smoothed trend as one possible nonlinear function of time $f(t)$ as above. See Section 4 for more details.

innovations. The stochastic volatility is $Var[\mathbf{v_t}|\mathbf{\Lambda_t}] = \mathbf{B}^{-1}\mathbf{\Lambda_t}\mathbf{B}^{-1'} = \mathbf{\Gamma_t}$, and conditionally standardized innovations $(\epsilon_\mathbf{t})$ are i.i.d. Gaussian, for any volatility history. Thus these standardized innovations are independent of the volatility process $\mathbf{\Gamma_t}$.

The dynamic of the volatility process is very constrained, since the serial dependence arises only through the diagonal matrix $\mathbf{\Lambda_t}$, not by means of $\mathbf{B}$, which is unchanging across time. Finally, the natural logarithms of the diagonal elements of the $\mathbf{\Lambda_t}$ matrix are assumed to follow independent Gaussian random walks.

A few points of discussion are worth mentioning. First, if $\lambda_{i,t} = \lambda_{i,t-1}$ for all t, then the underlying processes, $\lambda_{i,t}$, that drive the volatility of $\mathbf{v_t}$ cannot be identified independently of the $\mathbf{B}$ matrix. Moreover, the choice of $\mathbf{B}$ being constrained to be lower-triangular solves the identification problem of identifying the elements of $\mathbf{\Lambda_t}$ from those of $\mathbf{B}^{-1}$.

Note that the Clark specification is not invariant to permutations in the asset order within the VAR. Indeed, without loss of generality, let us consider the bivariate case. The variance of the innovation $\mathbf{v_t}$ is

$$E[\mathbf{v_t}\mathbf{v_t}'] = E[\mathbf{B}^{-1}\mathbf{\Lambda_t}^{0.5}\epsilon_\mathbf{t}\epsilon_\mathbf{t}'\mathbf{\Lambda_t}^{0.5}(\mathbf{B}^{-1})'] = \mathbf{B}^{-1}\mathbf{\Lambda_t}(\mathbf{B}^{-1})' = \mathbf{\Gamma_t} \tag{2a}$$

$$= \begin{bmatrix} b^{11} & 0 \\ b^{21} & b^{22} \end{bmatrix} \begin{bmatrix} \lambda_{1,t} & 0 \\ 0 & \lambda_{2,t} \end{bmatrix} \begin{bmatrix} b^{11} & b^{21} \\ 0 & b^{22} \end{bmatrix} \tag{2b}$$

$$= \begin{bmatrix} (b^{11})^2\lambda_{1,t} & b^{11}b^{21}\lambda_{1,t} \\ b^{11}b^{21}\lambda_{1,t} & (b^{21})^2\lambda_{1,t} + (b^{22})^2\lambda_{2,t} \end{bmatrix}, \tag{2c}$$

where b^{ij}, $i, j = 1, 2$ denote the elements of the $\mathbf{B}^{-1}$ matrix.

Therefore, the variance of the innovation of the first series and the covariance between the two innovations depend only on the process $\lambda_{1,t}$. However, the variance of

the second series depends on *both* processes $\lambda_{1,t}$ and $\lambda_{2,t}$. Therefore, shocks to the processes $\xi_{i,t}$ have asymmetric effects on the variances of the innovations $v_{1,t}$ and $v_{2,t}$. This asymmetry is chosen arbitrarily by our order of assignment of the series to the VAR.

In the $Clark$ specification, the volatility and covolatility processes are nonstationary. By the properties of the Gaussian random walk, we get

$$\ln(\lambda_{i,t})|\ln(\lambda_{i,0}) \sim N[\ln(\lambda_{i,0}), t\varphi_i]. \tag{3}$$

We deduce that

$$E[\lambda_{i,t}|\lambda_{i,0}] = \lambda_{i,0}\exp\left(\frac{t\varphi_i}{2}\right). \tag{4}$$

On average we get an exponential rate of explosion of the diagonal elements of the matrix $\mathbf{\Lambda_t}$. If $\varphi_i > \varphi_j$, say, the volatility of series i becomes, asymptotically, infinitely larger than the volatility of series j. And so as $t \to \infty$ we have that, conditional on past information, the process $\lambda_{i,t}$ is divergent (i.e. explosive). This result implies that forecasts of the VAR innovation covariance matrices will have explosive elements, which is not a suitable property of Clark's model.

However, all is not lost; a similar type of argument shows that if we respecify the $\lambda_{i,t}$ process as

$$\ln(\lambda_{i,t}) = z_{i,t} = \alpha + \beta z_{i,t-1} + \xi_{i,t}, \tag{5}$$

then for $|\beta| < 1$ the process $\lambda_{i,t}$ is now convergent with unconditional mean

$$E[\lambda_{i,t}] = \exp\{\frac{\alpha}{1-\beta} + \frac{\varphi_i}{2(1-\beta^2)}\}. \tag{6}$$

3.2 Alternative volatility process specification

As an alternative to the $Clark$ specification, we propose respecifying the volatility process to more easily account for spill-over effects in covariance across time, through the use of a multivariate Inverse Wishart specification.

While in the Clark specification the nonstationary $\lambda_{i,t}$ processes drive the covariance dynamics, $\mathbf{\Gamma_t}$, through the lower-triangular $\mathbf{B}$ matrix, the Inverse Wishart model specifies the dynamics of the latent covariance matrices directly. Fundamentally, the Inverse Wishart process implies *stationary* covariance matrix dynamics. Therefore, the Inverse Wishart process also allows us to formulate the covariance matrix process as autoregressive with a finite unconditional mean that exists under certain conditions defined below.

The Inverse Wishart Stochastic Volatility ($IWSV$) model, is given as follows:

$$\mathbf{v_t} = \mathbf{\Pi}\left(L\right)\left(\mathbf{y_t} - \mathbf{\Psi d_t}\right), \tag{1a}$$

$$\mathbf{v_t}|\underline{\mathbf{\Sigma}}_{\mathbf{t}}, \underline{\mathbf{y}}_{\mathbf{t}} \sim MVN_p(\mathbf{0}, \mathbf{\Sigma_t}), \tag{7a}$$

$$\mathbf{\Sigma_t}|\underline{\mathbf{\Sigma}}_{\mathbf{t-1}}, \underline{\mathbf{y}}_{\mathbf{t}} \sim IW_p(\nu, \mathbf{S_{t-1}}), \tag{7b}$$

$$\text{where} \quad \mathbf{S_{t-1}} = \left(\mathbf{CC}' + \sum_{k=1}^{K} \mathbf{A}_k \mathbf{\Sigma_{t-k}} \mathbf{A}_k'\right)(\nu - p - 1), \tag{7c}$$

where $\underline{\mathbf{y}}_{\mathbf{t}}$, for instance, denotes the set of current and lagged values of $\mathbf{y_t}$, and $IW_p(\nu, \mathbf{S})$, denotes the Inverse Wishart distribution with dimension p, degree of freedom (i.e. shape parameter) ν, and scale matrix $\mathbf{S}$.[13] The specification of the scale matrix in (7c) is the same as in the multivariate ARCH models considered in Engle and Kroner (1995). In particular, the $p \times p$ matrices, $\mathbf{C}$ and $\mathbf{A_k}$, $k = 1, \ldots, K$, are identified if $\mathbf{C}$ is lower

[13] A stochastic, symmetric, positive-definite matrix $\mathbf{\Sigma}$ follows the Inverse Wishart distribution: $\mathbf{\Sigma} \sim IW_p(\nu, \mathbf{S})$, if and only if $\mathbf{\Sigma}^{-1}$ follows the Wishart distribution: $\mathbf{\Sigma}^{-1} \sim W_p(\nu, \mathbf{S}^{-1})$.

triangular with strictly positive elements and if the top left element of each $\mathbf{A_k}$ are strictly positive.

The Inverse Wishart distribution is a continuous distribution for stochastic, symmetric, positive-definite matrices [see e.g. Press (1982)]. The joint density function of the Inverse Wishart distribution has the simple analytic expression given as:

$$f(\mathbf{\Sigma}; \nu, \mathbf{S}) = \frac{|det(\mathbf{S})|^{\nu/2} |det(\mathbf{\Sigma})|^{-\frac{\nu+p+1}{2}}}{2^{\frac{vp}{2}} \Gamma_p(\nu/2)} \exp\left[-\frac{1}{2} Tr\left(\mathbf{S}\mathbf{\Sigma}^{-1}\right)\right], \tag{8}$$

where $Tr(\cdot)$ denotes the trace operator and $\Gamma_p(\cdot)$, the multivariate Gamma function.

The dynamics of the stochastic volatility matrix given in (7c) do not involve lagged values of the series variable $(\mathbf{y_t})$. Thus, the stochastic volatility is exogenous and the $IWSV$ specification assumes no leverage effects.

From the properties of the Inverse Wishart distribution, we deduce the first and second-order conditional moments of the volatilities [Press (1982)]:

$$E[\mathbf{\Sigma_t} | \underline{\mathbf{\Sigma}}_{\mathbf{t-1}}] = \frac{\mathbf{S_{t-1}}}{\nu - p - 1} = \mathbf{C}\mathbf{C}' + \sum_{k=1}^{K} \mathbf{A}_k \mathbf{\Sigma_{t-k}} \mathbf{A}_k', \tag{9a}$$

$$Var[\sigma_{ij,t} | \underline{\mathbf{\Sigma}}_{\mathbf{t-1}}] = \frac{(\nu - p + 1)\, s_{ij,t-1}^2 + (\nu - p - 1)\, s_{ii,t-1} s_{jj,t-1}}{(\nu - p)(\nu - p - 1)^2 (\nu - p - 3)}, \tag{9b}$$

$$\text{and} \quad Cov[\sigma_{ij,t}, \sigma_{kl,t} | \underline{\mathbf{\Sigma}}_{\mathbf{t-1}}] = \frac{2 s_{ij,t} s_{kl,t} + (\nu - p - 1)\left(s_{ik,t} s_{jl,t} + s_{il,t} s_{kj,t}\right)}{(\nu - p)\left(\nu - p - 1\right)^2 (\nu - p - 3)}, \tag{9c}$$

where σ_{ij} is the ijth element of $\mathbf{\Sigma_t}$ and $s_{ij,t-1}$ is the ijth element of $\mathbf{S_{t-1}}$.

This specification is similar to that in Philipov and Glickman (2006) although we modify it slightly. For one, we add the constant matrix $\mathbf{C}\mathbf{C}'$ to the scale matrix expression (7c) in order to allow for a non-zero unconditional mean of the volatility process. Secondly, we add a number of lags K instead of just one. Finally, Philipov and

Glickman (2006) employ an extra parameter, d, to allow for a geometric autoregressive recursion of varying rates, as opposed to the fixed arithmitic average employed here. For instance, they consider the similar model with autoregressive lag order set to 1, but allow for the lagged effect to be taken account by means of a $\mathbf{\Sigma_{t-1}}^d$ matrix:[14]

$$\mathbf{S_{t-1}} = \nu \mathbf{A}^{-1/2'} \mathbf{\Sigma_{t-1}}^d \mathbf{A}^{-1/2} \tag{10}$$

Note, the purpose here is not to improve on the Philipov and Glickman model, but rather to suggest something similar as a useful alternative to Clark (2011) in terms of forecasting.

At this point we present weak stationarity conditions of the $IWSV$ volatility process.

Proposition 3.1. *Existence of the unconditional mean of the $IWSV$ process*

The unconditional mean of the $IWSV$ process exists if and only if all the eigenvalues of the matrix $\mathbf{\Upsilon} = \sum_{k=1}^{\mathrm{K}} \mathbf{\Xi_k}$ are less than 1 in modulus. In this case the unconditional mean is given by:

$$E\left[\sigma_t\right] = \left(\mathbf{I}_g - \mathbf{\Upsilon}\right)^{-1} c, \tag{11}$$

where $g = \frac{p(p+1)}{2}$, $c = vech\left(\mathbf{C}\mathbf{C}'\right)$, $\sigma_t = vech\left(\mathbf{\Sigma_t}\right)$, and $\mathbf{\Xi_i} = \mathbf{L}\left(\mathbf{A_i} \otimes \mathbf{A_i}\right)\mathbf{D}$. The existence of the unconditional mean is a necessary condition for (weak) stationarity.

[14] Note that in Philipov and Glickman (2006) the use of notation is different. For example, they have that $\mathbf{\Sigma_t}^{-1} | \nu, \mathbf{S}_{t-1} \sim W_p(\nu, \mathbf{S_{t-1}})$, where $\mathbf{S_{t-1}} = \frac{1}{\nu}\mathbf{A}^{1/2}\left(\mathbf{\Sigma_{t-1}}^{-1}\right)^d \mathbf{A}^{1/2'}$. Therefore, this implies that $\mathbf{\Sigma_t} | \nu, \mathbf{S}_{t-1}^{-1} \sim IW_p(\nu, \mathbf{S_{t-1}^{-1}})$ and so depends on the inverse scale matrix instead of the scale matrix. Therefore our scale matix is the inverse of theirs.

Note that in this case $\mathbf{L}$ and $\mathbf{D}$ are the elimination and duplication matrices respectively so that $vec\,(\mathbf{X}) = \mathbf{D}vech(\mathbf{X})$ and $vech\,(\mathbf{X}) = \mathbf{L}vec(\mathbf{X})$.[15]

Proof of Proposition 3.1

From equation (9a) we have:

$$\mathbf{\Sigma_t} = \mathbf{C}\mathbf{C}^{'} + \sum_{k=1}^{K} \mathbf{A_k}\mathbf{\Sigma_{t-k}}\mathbf{A_k}^{'} + \mathbf{Z_t}, \tag{12}$$

where $\mathbf{Z_t}$ is a mean zero matrix of weak white noises.

First, by recursive substitution of $\mathbf{\Sigma_{t-i}}$, $i = 1, \ldots$ we can show that the right hand side of (12) converges in expectation. Next, taking unconditional expectation $\hat{\mathbf{\Sigma}} = E[\mathbf{\Sigma_t}]$ we have:

$$\hat{\mathbf{\Sigma}} = \mathbf{C}\mathbf{C}^{'} + \sum_{k=1}^{K} \mathbf{A_k}\hat{\mathbf{\Sigma}}\mathbf{A_k}^{'}. \tag{13}$$

Vectorizing above we have:

$$vec(\hat{\mathbf{\Sigma}}) = vec(\mathbf{C}\mathbf{C}^{'}) + \sum_{k=1}^{K} vec(\mathbf{A_k}\hat{\mathbf{\Sigma}}\mathbf{A_k}^{'}) \tag{14a}$$

$$= vec(\mathbf{C}\mathbf{C}^{'}) + \sum_{k=1}^{K} \left(\mathbf{A_k} \otimes \mathbf{A_k}^{'}\right) vec(\hat{\mathbf{\Sigma}}) \tag{14b}$$

$$\Rightarrow \mathbf{L}vec(\hat{\mathbf{\Sigma}}) = \mathbf{L}vec(\mathbf{C}\mathbf{C}^{'}) + \sum_{k=1}^{K} \mathbf{L}\left(\mathbf{A_k} \otimes \mathbf{A_k}^{'}\right) \mathbf{D}vech(\hat{\mathbf{\Sigma}}) \tag{14c}$$

$$\Rightarrow vech(\hat{\mathbf{\Sigma}}) = vech(\mathbf{C}\mathbf{C}^{'}) + \sum_{k=1}^{K} \mathbf{\Xi_k} vech(\hat{\mathbf{\Sigma}}). \tag{14d}$$

[15]The duplication matrix is the unique $n^2 \times n(n+1)/2$ matrix, $\mathbf{D}$, which, for any $n \times n$ symmetric matrix $\mathbf{X}$, transforms $vech(\mathbf{X})$ into $vec(\mathbf{X})$, where $vec(\cdot)$ is the vectorization operator which maps from the $n \times n$ dimensional space to the $n^2 \times 1$ dimensional space and $vech(\cdot)$ is the operator that omits the lower (resp. upper) triangle of the symmetric matrix $\mathbf{X}$ so that it maps from the $n \times n$ dimensional space into the $n(n+1)/2 \times 1$ dimensional space. The elimination matrix performs the inverse operation: it is the unique $n(n+1)/2 \times n^2$ matrix, $\mathbf{L}$, which, for any $n \times n$ symmetric matrix $\mathbf{X}$, transforms $vec(\mathbf{X})$ into $vech(\mathbf{X})$. See Magnus and Neudecker (1980) for more details.

And so Proposition 3.1 follows.

The condition in Proposition 3.1 is a necessary condition for stationarity, but not a sufficient condition. In a Bayesian approach, we are interested in the whole distribution, not in the mean only. Thus strict stationarity, that is concerning the entire distribution, has to be considered, not only weak stationarity. Unfortunately, necessary and sufficient conditions for the strict stationarity of the autoregressive Inverse Wishart process have not yet been derived in the literature.[16]

3.3 Comments

The advantages of such a change to the specification of the volatility process defined by the $IWSV$ model are as follows:

1. The direct specification of the dynamics of the latent stochastic volatility process, $\mathbf{\Sigma_t}$, precludes the need to specify a $\mathbf{B}$ matrix.

2. These autoregressive dynamics between volatility series are more easily interpreted as volatility spill-over effects, since we no longer need to disentangle the convoluted relationships implied by the $\mathbf{B}$ matrix and the independent volatility driving processes, $\lambda_{i,t}$.

3. The model is now invariant to permutation of the order of the observed series.

4. As was shown, it is easy to derive conditions ensuring the existence of the unconditional mean of the processes $(\mathbf{\Sigma_t})$ and $(\mathbf{y_t})$. However, this condition is a weak one and the condition of strong stationarity remains to be shown.

[16] Whereas they have been derived for the analogue Wishart autoregressive process (WAR) [see Gourieroux et al. (2009)].

5. The assumption of stationary volatility dynamics resolves a problem with forecasts, since assuming nonstationary volatility would make our forecast density prediction intervals "blow up" as the horizon increases.

In both the $Clark$ and $IWSV$ specifications, the volatility processes are coupled to the vector autoregressive process, with trends, for the observed variables $\mathbf{y_t}$. Both specifications have a state-space representation with the $VAR(J)$ as observation equation and the $IWSV$ or $Clark$ as the state equation, the covariance being the "state" of the model. In control engineering, a state-space representation is a mathematical model of a physical system as a set of input, output and state variables related by first-order differential equations. However, in this case the model is not a linear state-space representation [see the system in (7)]. Therefore, the standard Kalman Filter algorithm for extracting the state from the noise will not be optimal.

Note that our Inverse Wishart model can be inverted to show the precision matrix as Wishart distributed (omitting the $\mathbf{CC}'$ constant matrix for simplicity) as

$$\mathbf{\Sigma_t}^{-1}|\underline{\mathbf{\Sigma_{t-1}}} \sim W_p(\nu, \left(\sum_{k=1}^{K} \mathbf{A_k \Sigma_t A_k}'\right)^{-1}), \tag{15}$$

or after a change of notation $\mathbf{\Omega_t} = \mathbf{\Sigma_t}^{-1}$ as

$$\mathbf{\Omega_t}|\underline{\mathbf{\Omega_{t-1}}} \sim W_p(\nu, \left(\sum_{k=1}^{K} \mathbf{A_k \Omega_t}^{-1} \mathbf{A_k}'\right)^{-1}). \tag{16}$$

However, an autoregressive Wishart matrix process, $\mathbf{\Omega_t}$, say, is usually written as [Gourieroux et al. (2009)]

$$\mathbf{\Omega_t}|\underline{\mathbf{\Omega_{t-1}}} \sim W_p(\nu, \left(\sum_{k=1}^{K} \mathbf{A_k \Omega_t A_k}'\right)). \tag{17}$$

And so, the only difference between the two models is the specification of the scale matrix: as an arithmetic average in the case of the standard Wishart autoregressive process, or as a harmonic average in our case. Therefore, when the lag order of the autoregression $K > 1$, we have an asymmetry between the behaviour of the precision matrix and covariance matrix processes. Of course, this issue does not affect Philipov and Glickman (2006) since their autoregression is only of order 1.

Since we have that

$$E\left[\frac{1}{x}\right] \geq \frac{1}{E[x]} \quad \Leftrightarrow \left(E\left[\frac{1}{x}\right]\right)^{-1} \leq E[x], \tag{18}$$

we can expect that the harmonic average is smaller than an arithmetic average (even in the case of matrices, but we omit the proof). It is possible that this inequality maybe useful in deriving sufficient conditions for strict stationary for the $IWSV$ autoregressive model.

As an aside, Fox and West (2011) also propose a novel class of stationary covariance matrix processes which exploit properties of Inverse Wishart partitioned matrix theory. Specifically, by augmenting the parameter state-space they show that we can easily obtain representations for the terms in a factorization of the joint density of covariance matrices across time, $f(\mathbf{\Sigma}_T, \ldots, \mathbf{\Sigma}_0) = \prod_{t=1}^{T} f(\mathbf{\Sigma}_t, \mathbf{\Sigma}_{t-1}) \prod_{t=2}^{T} f(\mathbf{\Sigma}_{t-1})$. This expression defines a stationary first-order Markov process on the covariance matrices across time, with the marginal distribution given as $\mathbf{\Sigma}_t \sim IW_q(\nu + 2, \nu\mathbf{S})$, and given the following augmented matrix

$$\begin{pmatrix} \mathbf{\Sigma}_t & \phi_t^{'} \\ \phi_t & \mathbf{\Sigma}_{t-1} \end{pmatrix} \sim IW_{2q}\left(\nu + 2, \nu \begin{pmatrix} \mathbf{S} & \mathbf{SF}^{'} \\ \mathbf{FS} & \mathbf{S} \end{pmatrix}\right), \tag{19}$$

where $\phi_t = \boldsymbol{\Upsilon}_t \boldsymbol{\Sigma}_{t-1}$, we have that by Inverse Wishart partitioned matrix theory the covariance process can be written as an AR(1) process $\boldsymbol{\Sigma}_t = \boldsymbol{\Psi}_t + \boldsymbol{\Upsilon}_t \boldsymbol{\Sigma}_{t-1} \boldsymbol{\Upsilon}_t^{'}$, with $\boldsymbol{\Upsilon}_t$ representing a random coefficient matrix and $\boldsymbol{\Psi}_t$ representing an innovation (note both $\boldsymbol{\Upsilon}_t$ and $\boldsymbol{\Psi}_t$ are latent variables). Under this framework the conditional density $\boldsymbol{\Sigma}_t | \boldsymbol{\Sigma}_{t-1}$ is not of an analytical form but can nonetheless be explored theoretically. See Fox and West (2011) for more details.

4 Priors

The Bayesian estimation framework employed requires of us to specify certain prior beliefs upon the parameter set and this is done through the specification of prior densities. In most cases the prior densities are chosen to be conditionally conjugate—that is, they are chosen of a known family such that the conditional posterior density, i.e. the density of a particular parameter, conditional on both the other parameters and the data, works out to be of the same family as the prior. This facilitates estimation greatly since the need for arbitrarily choosing a suitable proposal density, as in a Metropolis-Hastings algorithm (MH), is avoided completely—in fact, in this case the proposal is always accepted and the MH algorithm is just a special case of the Gibbs sampler [Greenberg (2008), pg.101]. The following Sections outline the specific families the prior densities take, as well as chosen values for hyperparameters.

$Clark$ and $IWSV$ specifications share the same dynamic model for the observed macroeconomic time series, $\mathbf{y_t}$, given the volatility path, with parameters $\theta_1 = \{\boldsymbol{\Psi}, \boldsymbol{\Pi_1}, \ldots, \boldsymbol{\Pi_J}\}$, but different dynamics for the volatility process, with parameters $\theta_2 = \{\mathbf{B}, \boldsymbol{\Phi}\}$ for the $Clark$ model[17] and $\theta_2 = \{\mathbf{C}, \nu, \mathbf{A_1}, \ldots, \mathbf{A_K}\}$ for the $IWSV$

[17]Where $\boldsymbol{\Phi}$ is the diagonal matrix with the variances of the $\lambda_{i,t}$ volatility driving process shocks, φ_i,

model. We assume that parameters θ_1 and θ_2 are independent under the prior distributions. We describe below in greater detail the priors for both θ_1 and θ_2.

4.1 VAR(J) priors

i) Prior on Π

The prior for the VAR coefficients $\mathbf{\Pi_j}$, follow a modified *Minnesota specification* (see Litterman (1986)). In this case we assume that the prior for the joint density of $\mathbf{\Pi}' = [\mathbf{\Pi_1}, \mathbf{\Pi_2}, \ldots, \mathbf{\Pi_J}]$ is Normal, $\mathbf{\Pi} \sim N[\mu_{\mathbf{\Pi}}, \mathbf{\Xi_\Pi}]$, where the autoregressive order J is assumed known. Moreover, the prior mean of the joint density of the elements of $\mathbf{\Pi}$ assumes that the VAR follows an AR(1) process, i.e. means of the prior density for all the elements of autoregressive matrices beyond lag 1 are set to 0. Since GDP growth displays more autoregressive decay in levels, we set its first-order autoregressive prior mean to 0.25 and set the others to 0.8. Cross equation prior means, that is, the means for the prior density of the off-diagonal elements of $\mathbf{\Pi_{ik,1}}$ for $i \neq k$, are also set to 0.

Let us now explain how the variances and covariances of the prior are chosen. Minnesota "own equation" variances, that are, the variances of the prior density for the main diagonal elements of $\mathbf{\Pi_j}$, shrink as a harmonic series for each additional lag (i.e. $\omega_{ii,j} = \frac{0.2}{j}$ for $j = 1, \ldots J$) . Also, "cross equation" variances are typically set to $\omega_{ik,j} = 0.5(\frac{0.2}{j} \times \frac{\sigma_i^*}{\sigma_k^*})$, where σ_i^* is the estimated standard error of the residuals from a univariate autoregression on the ith macroeconomic series with six lags, pre-fit to the data in advance. For simplicity, however, we will employ $\omega_{ik,j} = 0.5(\frac{0.2}{j})$ instead, that is, the variance of the prior density for the i,kth element of $\mathbf{\Pi_j}$.

along its main diagonal.

ii) Prior on $\mathbf{\Psi}$

Priors on the deterministic parameters of $\mathbf{\Psi}$ defining the trend are extremely important given the modest sample size employed and are chosen as to influence the series' trends toward certain reasonable values. In the case where the trend is assumed constant: $\mathbf{d_t} = 1$, and $\mathbf{\Psi}$ is a vector. This dramatically reduces the number of parameters that need to be estimated as the number of series increases. However, it places a prior constraint on the model by assuming that the selected constant trend chosen is correct. On the other hand, if we allow the trends to enter individually through the $\mathbf{d_t}$ term where $\mathbf{d_t}$=$[1, f(t-1), g(t)]'$, and $f(t)$ and $g(t)$ are exponentially smoothed trends for the unemployment rate and inflation growth respectively, then $\mathbf{\Psi}$ becomes a $p \times 3$ matrix from which we can statistically evaluate whether the relevant diagonal elements are indeed equal to 1 (which would imply the trends are in fact correct).

In either case we assume that the prior for the joint density of the elements of $\mathbf{\Psi}$ is Normal, $\mathbf{\Psi} \sim N[\mu_{\mathbf{\Psi}}, \mathbf{\Xi_\Psi}]$. Moreover, we assume that the priors for $\mathbf{\Psi}$ and $\mathbf{\Pi}$ are independent. GDP growth is influenced to have a constant trend around 3.0% through its prior mean, while inflation and unemployment are influenced to center around their trends, $g(t)$ and $f(t-1)$ which are exponentially smoothed values of inflation growth and the unemployment rate respectively (see Section 2). Finally, the interest rate is centered around the same trend as inflation; however, we also add to this the constant trend of 2.5% to reflect the real long-run rate. More precisely, for the macroeconomic series taken in order as: GDP growth, the inflation rate, the interest rate, and unemployment

rate, the prior mean of $\mathbf{\Psi}$ will take the form

$$\begin{bmatrix} 3.0 \\ 0 \\ 2.5 \\ 0 \end{bmatrix} \tag{20}$$

when the trends are constant, and

$$\begin{bmatrix} 3.0 & 0 & 0 \\ 0 & 0 & 1 \\ 2.5 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{21}$$

when the trends are driven by the 3-dimensional $\mathbf{d_t}$.

The prior variances of $\mathbf{\Psi}$ are set as follows: GDP growth, 0.2 (0.3); inflation, 0.2 (0.3); the interest rate, 0.6 (0.75); and unemployment, 0.2 (0.3)—where these values have been adopted directly from Clark (2011). The first values, not in parenthesis, represent those employed in the recursive estimation scheme and are tighter since the gradually increasing sample size tends to limit the influence of the prior. Prior covariances for the elements of $\mathbf{\Psi}$ are set to zero.

4.2 Volatility model priors

i) Clark model

For the Clark (2011) model, priors on the volatility components of the model are as follows. The prior density for the elements of $\mathbf{B}$ is multivariate Normal and the prior

for each of the φ_i, $i = 1, \ldots, p$ is Inverse Gamma, and under the prior, the elements of $\mathbf{B}$ and the φ_i, $i = 1, \ldots, p$, are independent. Finally, we borrow numerical values of the hyperparameters directly from Clark's paper [Clark (2011) pg.331].

ii) IWSV model

For the Inverse Wishart autoregressive specification, we employ independent multivariate Normal priors on both $\mathbf{A_k}$, $\forall k$ and $\mathbf{C}$, and independently, a Gamma prior on $(\nu - p)$. The Gamma prior is set with hyperparameters $\alpha = 30, \beta = 2$ (shape and rate) as to represent ignorance of its value while the multivariate Normal priors for the $\mathbf{C}$ and $\mathbf{A_k}$'s are set somewhat loosely to let the data speak. In this respect, prior means for the main diagonal of $\mathbf{C}$ are 0.3 and the main diagonal of $\mathbf{A_1}$ is set to 0.9 (both prior densities for the off-diagonals elements have zero means, and the means of all other elements of the $\mathbf{A_k}$, $k = 1, \ldots, K$ matrices are set to 0). Variances are set equal to 0.002 (i.e. standard deviation of about 0.045), and all covariances are set to zero.

5 Model estimation

Both $Clark$ and $IWSV$ model specifications are estimated within the Bayesian framework using Markov Chain Monte-Carlo (MCMC) techniques, particularly the Gibbs sampler.

i) Gibbs sampler

Indeed, by selecting prior distributions in conjugate families, we can derive closed form expressions of conditional posterior distributions. For expository purposes, let us consider a case in which the set of parameters can be divided into two subsets, θ_1 and θ_2, such that we know the expression of conditional posterior distributions $p(\theta_1|\theta_2, y)$ and $p(\theta_2|\theta_1, y)$. Let us also assume that it is easy to draw in these conditional posterior

distributions. In general, it is not possible to obtain the closed form expression for the joint posterior distribution $p(\theta_1, \theta_2|y)$.

The Gibbs sampler is a method to derive numerically a good approximation of the joint posterior, while also allowing us to draw in this posterior. The idea is to consider the Markov process $\theta^{(m)}$, defined recursively by:

1. $\theta_1^{(m)}$ is drawn in $p(\theta_1|\theta_2^{(m-1)}, y)$,

2. $\theta_2^{(m)}$ is drawn in $p(\theta_2|\theta_1^{(m)}, y)$.

For large $m \geq M$, the values $\theta^{(m)}$ will approximately follow the invariant distribution of the Markov process $\theta^{(m)}$, that is, the joint posterior. In particular, $\theta^{(m)}$, for large m, is a drawing in $p(\theta_1, \theta_2|y)$.

This approach is easily extended when the set of parameters is divided into more than two subsets [see below the sequence used for both the $Clark$ and $IWSV$ specifications].

ii) Augmented parameters

In a Bayesian framework there is little difference between parameter θ and latent volatilities $\mathbf{\Sigma_t}$, $t = 1, \ldots, T$. They are both unobserved and stochastic. Therefore, the Gibbs sampler can be applied jointly to θ and $\underline{\mathbf{\Sigma_T}} = \{\mathbf{\Sigma_1}, \ldots, \mathbf{\Sigma_T}\}$ to reconstitute the joint density $p(\theta, \underline{\mathbf{\Sigma_T}}|\underline{\mathbf{y_T}})$. This joint density has two components, that are $p(\theta|\underline{\mathbf{y_T}})$, which is the posterior distribution of the parameter, and $p(\underline{\mathbf{\Sigma_T}}|\theta, \underline{\mathbf{y_T}})$, which is the filtering distribution of the sequence of latent volatilities.

iii) Gibbs sampler steps - Clark (2011)

Specifically, given the parameters described in the Section 3.1 above, we have the following Gibbs sampling steps for the Clark (2011) benchmark model, where the volatility driving process $\underline{\mathbf{\Lambda_T}}$ is introduced as an augmented parameter to be estimated.

All conditional posteriors below are conditional on the data, $\underline{\mathbf{y_T}}$, left unstated below for ease of exposition.

1. Draw the autoregressive coefficients $\mathbf{\Pi}' = [\mathbf{\Pi_1}, \mathbf{\Pi_2}, \ldots, \mathbf{\Pi_J}]$ of the VAR, conditional on $\mathbf{\Psi}$, $\underline{\mathbf{\Lambda_T}}$, $\mathbf{B}$, and $\mathbf{\Phi}$=$diag(\varphi_1, \varphi_2, \ldots, \varphi_p)$, given a conditionally conjugate multivariate Normal prior, $\mathbf{\Pi} \sim N(\mu_{\mathbf{\Pi}}, \mathbf{\Xi_\Pi})$.

2. Draw the trend coefficients $\mathbf{\Psi}$ of the VAR, conditional on $\mathbf{\Pi}$, $\underline{\mathbf{\Lambda_T}}$, $\mathbf{B}$, and $\mathbf{\Phi}$=$diag(\varphi_1, \varphi_2, \ldots, \varphi_p)$, given a conditionally conjugate multivariate Normal prior, $\mathbf{\Psi} \sim N(\mu_{\mathbf{\Psi}}, \mathbf{\Xi_\Psi})$.

3. Draw the elements of $\mathbf{B}$ (lower triangular with ones in the diagonal) conditional on $\mathbf{\Pi}$, $\mathbf{\Psi}$, $\underline{\mathbf{\Lambda_T}}$, and $\mathbf{\Phi}$=$diag(\varphi_1, \varphi_2, \ldots, \varphi_p)$, given Normal, independent, priors on each of the elements of the $\mathbf{B}$ matrix.

4. Draw the elements of the volatility driving process $\mathbf{\Lambda_t}$ for each time $t = 1, \ldots, T$ in sequence, each conditional on $\mathbf{\Lambda_{\backslash t}}, \mathbf{\Pi}$, $\mathbf{\Psi}$, $\mathbf{B}$, and $\mathbf{\Phi}$=$diag(\varphi_1, \varphi_2, \ldots, \varphi_p)$, where the notation $\backslash \mathbf{t}$ denotes the set of all matrices except that at time t. A **Metropolis-Hastings-within-Gibbs step** is required here since the posterior distribution is of an unknown family.

5. Draw the diagonal elements of $\mathbf{\Phi}$ conditional on $\mathbf{\Pi}$,, $\mathbf{\Psi}$, $\mathbf{B}$, and $\underline{\mathbf{\Lambda_T}}$. We assume conditionally conjugate, independent Inverse-Gamma priors on each $\varphi_i \sim IG(\frac{\gamma}{2}, \frac{\delta}{2})$.

iv) Gibbs sampler steps - IWSV model

Similarly, we have the following Gibbs sampling steps for the $IWSV$ specification, where the covariance matrices $\mathbf{\Sigma_t}$ are introduced as augmented variables. Again, all conditional posteriors below are implicitly conditional on the data, $\underline{\mathbf{y_T}}$.

1. Draw the slope coefficients $\mathbf{\Pi}^{'} = [\mathbf{\Pi_1}, \mathbf{\Pi_2}, \ldots, \mathbf{\Pi_J}]$ of the VAR, conditional on $\mathbf{\Psi}$, $\underline{\mathbf{\Sigma_T}}$, $\mathbf{A}_k, \forall k$, $\mathbf{C}$, and ν, given multivariate Normal prior, $\mathbf{\Pi} \sim N(\mu_{\mathbf{\Pi}}, \mathbf{\Xi_\Pi})$.

2. Draw the steady state coefficients $\mathbf{\Psi}$ of the VAR, conditional on $\mathbf{\Pi}$, $\underline{\mathbf{\Sigma_T}}$, $\mathbf{A}_k, \forall k$, $\mathbf{C}$, and ν, given multivariate Normal prior, $\mathbf{\Psi} \sim N(\mu_{\mathbf{\Psi}}, \mathbf{\Xi_\Psi})$.

3. Draw the parameters $\mathbf{A}_k, \ \forall k, \mathbf{C}$, and ν jointly, conditional on $\mathbf{\Pi}$, $\mathbf{\Psi}$, and $\underline{\mathbf{\Sigma_T}}$. Multivariate Normal priors are imposed on the $\mathbf{A}_k$ matrices, and the $\mathbf{C}$ matrix. A Gamma prior is imposed on the degree of freedom, ν. A **Metropolis-Hastings-within-Gibbs step** is required here since the posterior distribution is of an unknown family.

4. Finally, draw $\mathbf{\Sigma_t}$ conditional on $\mathbf{\Sigma}_{\backslash \mathbf{t}}, \mathbf{A}_k, \ \forall k, \mathbf{C}, \ \nu, \ \mathbf{\Pi}$, and $\mathbf{\Psi}$, $\forall t$ in sequence. A **Metropolis-Hastings-within-Gibbs step** is required here since the posterior distribution is of an unknown family.

The steps above are derived in greater detail within Appendix 11.

Under the $IWSV$ process, the direct estimation of the latent stochastic volatility covariance matrix process increases the number of latent parameters from Tp to $Tp(p+1)/2$, raising quickly the curse of dimensionality as an issue as p increases. Moreover, the number of regular parameters goes from $\frac{p(p-1)}{2} + p$ to $\frac{p(p+1)}{2} + Kp^2 + 1$, although in this latter case many possible reparameterizations are possible to reduce the number of regular parameters the $IWSV$ must estimate.

Moreover, since conditionally conjugate priors are unknown at this point for the conditional posterior densities of the $IWSV$ regular parameters, a Metropolis-Hastings random walk sampler is employed. This additional Metropolis-Within-Gibbs step requires some extra work to obtain reasonable draws.

Finally, the estimation methods employed for both specifications suffer from the fact that they draw the latent stochastic volatility covariance matrices sequentially across time rather than jointly. Clearly, if the volatilities $\mathbf{\Sigma_t}$ are significantly correlated across time, joint sampling would prove superior. This is since the conditional posterior density may suffer from low variance, leading to draws which may fail to traverse the full parameter support in a reasonable amount of time. See Greenburg (2008) pg.94 for a simple example illustrating the problem.

6 Forecasts

6.1 Point and interval forecasts

Given the Bayesian model estimation framework employed, forecasts can be easily obtained with little extra computational overhead. Moreover, the Bayesian framework provides an intuitive way of comparing forecast accuracy.

Generally, the desired predictive density of some forecasted value y_f given data set y is

$$p\left(y_f \mid y\right) = \int p\left(y_f \mid \theta, y\right) \pi\left(\theta \mid y\right) d\theta \tag{22}$$

where $p(y_f|\theta, y)$ is the predictive distribution given parameter θ and data y, and $\pi(\theta|y)$ is the posterior distribution of θ. When the value y_f represents the true outcome of the data series known *ex post*, given the particular model formulation estimated *ex ante*, the left hand side is known as the *predictive likelihood* of the given outcome value y_f [Geweke and Amisano (2010)].

In our specification, formula (22) can be applied to forecast a future path $\mathbf{y_{T+1}}, \ldots, \mathbf{y_{T+h}}$

at date T. Moreover, we can introduce explicitly the stochastic volatility. We get:

$$p(\mathbf{y_{T+1}}, \ldots, \mathbf{y_{T+H}} | \underline{\mathbf{y_T}}) = \int \cdots \int p(\mathbf{y_{T+1}}, \ldots, \mathbf{y_{T+H}}, \mathbf{\Sigma_{T+1}}, \ldots, \mathbf{\Sigma_{T+H}}, \theta, \underline{\mathbf{\Sigma_T}} | \underline{\mathbf{y_T}}) \cdot d\theta d\underline{\mathbf{\Sigma_T}} \prod_{h=1}^{H} d\mathbf{\Sigma_{T+h}} \quad (23)$$

where again, $\underline{\mathbf{y_T}} = \{\mathbf{y_T}, \ldots, \mathbf{y_0}\}$. By factorizing the joint density within the integral in (23), we get

$$p(\mathbf{y_{T+1}}, \ldots, \mathbf{y_{T+H}} | \underline{\mathbf{y_T}}) = \int \cdots \int p(\mathbf{y_{T+1}}, \ldots, \mathbf{y_{T+H}} | \underline{\mathbf{\Sigma_{T+H}}}, \underline{\mathbf{y_T}}, \theta) \cdot p(\mathbf{\Sigma_{T+1}}, \ldots, \mathbf{\Sigma_{T+H}} | \underline{\mathbf{\Sigma_T}}, \underline{\mathbf{y_T}}, \theta) p(\underline{\mathbf{\Sigma_T}}, \theta | \underline{\mathbf{y_T}}) d\theta d\underline{\mathbf{\Sigma_T}} \prod_{h=1}^{H} d\mathbf{\Sigma_{T+h}}. \quad (24)$$

We already possess draws from $p(\underline{\mathbf{\Sigma_T}}, \theta | \underline{\mathbf{y_T}}) = p(\underline{\mathbf{\Sigma_T}} | \theta, \underline{\mathbf{y_T}}) p(\theta | \underline{\mathbf{y_T}})$ by using the Gibbs sampler with augmented parameter. Therefore, each time we draw an mth value from the Gibbs sampler for $\theta^{(m)}$ and $\underline{\mathbf{\Sigma_T}}^{(m)}$, we can simultaneously draw sequences $\{\mathbf{y_T}^{(m)}, \ldots, \mathbf{y_{T+H}}^{(m)}\}$ and $\{\mathbf{\Sigma_T}^{(m)}, \ldots, \mathbf{\Sigma_{T+H}}^{(m)}\}$ given the parameterization of the model and its implied conditional normality.

To summarize, we have the following steps:

1. Obtain a draw for $\theta^{(m)}$ and $\underline{\mathbf{\Sigma_T}}^{(m)}$ from the Gibbs sampler.

2. Conditioning on these values and the data, draw a covariance matrix $\mathbf{\Sigma_{T+1}}^{(m)}$ using the chosen volatility specification, either $Clark$, or $IWSV$.

3. Draw $\mathbf{y_{T+1}}^{(m)}$ from the $VAR(J)$ Gaussian level process.

4. Repeat again from step (1), drawing the next horizon, until we finish drawing for horizon $T + H$.

These steps above provide a draw from the joint conditional density

$$p(\mathbf{y_{T+1}}, \ldots, \mathbf{y_{T+H}}, \mathbf{\Sigma_{T+1}}, \ldots, \mathbf{\Sigma_{T+H}}, \theta, \underline{\mathbf{\Sigma_T}} | \underline{\mathbf{y_T}}). \tag{25}$$

These steps are repeated providing a sequence of draws for m sufficiently large, $M_0 + M > m \geq M_0$, say, with M_0 and M large.

The sequence of draws can be used for approximating the joint predictive distribution, $p(\mathbf{y_{T+1}}, \ldots, \mathbf{y_{T+H}} | \underline{\mathbf{y_{T+1}}})$, or some of its moments. Let us focus on the point and interval forecasts.

For the purposes of forecasting, we wish to consider both the conditional mean and quantiles of the predictive density

$$p(\mathbf{y_{T+1}}, \ldots, \mathbf{y_{T+H}} | \underline{\mathbf{y_T}}) = p(\mathbf{y_{T+H}} | \underline{\mathbf{y_{T+H-1}}}) \ldots p(\mathbf{y_{T+1}} | \underline{\mathbf{y_T}}). \tag{26}$$

i) Point forecasts

For the short-term horizon a consistent estimator of the mean of $E[\mathbf{y_{T+1}} | \underline{\mathbf{y_T}}]$ is its sample counterpart

$$\frac{1}{M} \sum_{m=M_0}^{M_0+M} \mathbf{y_{T+1}}^{(m)}, \tag{27}$$

computed on the final iteration of the Gibbs sampler.

Moreover, by the Law of Iterated Expectations, we have:

$$E[\mathbf{y_{T+1}} | \underline{\mathbf{y_T}}] = E[E[\mathbf{y_{T+1}} | \underline{\mathbf{\Sigma_T}}, \underline{\mathbf{y_T}}, \theta] | \underline{\mathbf{y_T}}]. \tag{28}$$

Therefore, another consistent estimator of $E[\mathbf{y_{T+1}}|\underline{\mathbf{y_T}}]$ is

$$\frac{1}{M}\sum_{m=M_0}^{M_0+M} E[\mathbf{y_{T+1}}|\underline{\mathbf{\Sigma_T}}^{(m)}, \underline{\mathbf{y_T}}, \theta^{(m)}], \tag{29}$$

as long as the conditional expectation $E[\mathbf{y_{T+1}}|\underline{\mathbf{\Sigma_T}}, \underline{\mathbf{y_T}}, \theta]$ has an analytical form.

Moreover, again by the Iterated Law of Expectations, the results above can be generalized to any horizon $h = 1, 2, \ldots, H$ since:

$$E[E[E[\mathbf{y_{T+h}}|\mathbf{y_{T+h-1}}, \ldots, \mathbf{y_{T+1}}, \underline{\mathbf{y_T}}]|\mathbf{y_{T+h-2}}, \ldots, \mathbf{y_{T+1}}, \underline{\mathbf{y_T}}]\ldots] = E[\mathbf{y_{T+h}}|\underline{\mathbf{y_T}}]. \tag{30}$$

ii) Interval forecasts

Interval forecasts at horizon h can be derived by estimating the quantiles of the predictive density $p(\mathbf{y_{T+h}}|\underline{\mathbf{y_T}})$. Let us look for a prediction interval with lower bound a α-quantile and upper bound the $(1-\alpha)$-quantile. This forecast interval can be approximated as follows:

1. Rank in increasing order the $\mathbf{y_{T+h}}^{(m)}$, $m = M_0 + 1, \ldots, M$.

2. The approximated confidence interval admits as lower bound the $\mathbf{y_{T+h}}^{(m)}$ at rank αM and an upper bound as the $\mathbf{y_{T+h}}^{(m)}$ at rank $(1-\alpha)M$.

6.2 Forecast comparison

i) Sample windows

The focus of this paper is to compare forecast performance. Forecast comparisons are made by employing a limited subsample of the data for estimation purposes, thus leaving some latter part of the data available as the "true" outcome of the macroeco-

nomic series. Furthermore, we do not simply estimate the parameters of the model once, rather we estimate the parameters a number of times in sequence, $n = 1, \ldots, N$, which we call "sample windows," each sample window focusing on a different subsample of the entire data set.

As in Clark (2011) we employ both "recursive" and "rolling" schemes for these sample windows. Under each scheme, sample windows are iterated upon: first, a subsample of the entire data set is isolated; we then use this subsample to estimate the model parameters, forecasts are then generated, and comparisons are made according to the chosen metrics discussed in this section. Both schemes employ the same subsample of the data for the first sample window. However, the two schemes differ in how they deal with later sample windows after the first. Under the recursive scheme, for each subsequent sample window, one future data point is appended to the end of the subsample, and so the subsample grows larger as the sequence of sample windows progress. Under the rolling scheme, the size of the subsample is fixed, and so the subsample shifts forward in time by one data point for each iteration.

The following **Figures 1** and **Figures 2** illustrate both the recursive and rolling window schemes where we have arbitrarily chosen a subsample size of 130 data points for the initial sample window. Notice how the time index of the last value in the subsample, T, changes across the iterations of sample windows, $n = 1, \ldots, N$, and so we can denote them $T(n)$ to show that they depend on the value of n.

ii) Comparing point forecasts

In comparing point forecasts we can employ the mean squared error (MSE) estimator for any forecast horizon $h = 1, \ldots, H$. Forecast errors can be computed where $\mathbf{y}^*_{\mathbf{T+h}}$ is the "true" out-of-sample data point at forecast horizon h and the point forecast $E[\mathbf{y_{T+h}}|\underline{\mathbf{y_T}}]$ is estimated as described above.

The mean squared error estimator is given as

$$MSE_{h,N} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{u_{T(n)+h}}\mathbf{u_{T(n)+h}}' \tag{31}$$

where the expected forecast error $E[\mathbf{y_{T(n)+h}}|\underline{\mathbf{y_{T(n)}}}] - \mathbf{y^*_{T(n)+h}}$ is computed for each iteration, n, and where $T(n)$ denotes the end of sample data point. In the case of the rolling window scheme, the set $\underline{\mathbf{y_{T(n)}}}$ denotes only those data points in the rolling window and not the entire set of data starting from $\mathbf{y_1}$.

The off-diagonals of $MSE_{h,N}$ represent the squared forecast errors across macro series at horizon h, and the main diagonal elements are the squared forecast errors for each individual series themselves.

iii) Model fit

The standard Bayesian tool for measuring the overall forecast performance is the *predictive likelihood* described above in (22) [Geweke and Amisano (2010)]. The finite sample approximation, within the context of our model, is given as:

$$\hat{p}(\mathbf{y^*_{T(n)+h}}, \ldots, \mathbf{y^*_{T(n)+1}}|\underline{\mathbf{y_{T(n)}}}) = \frac{1}{M}\sum_{m=M_0}^{M_0+M} p\left(\mathbf{y^*_{T(n)+h}}, \ldots, \mathbf{y^*_{T(n)+1}}|\underline{\mathbf{\Sigma_{T(n)}}}^{(m)}, \underline{\mathbf{y_{T(n)}}}, \theta^{(m)}\right)$$

$$\approx E\left[p\left(\mathbf{y^*_{T(n)+h}}, \ldots, \mathbf{y^*_{T(n)+1}}|\underline{\mathbf{\Sigma_{T(n)}}}, \underline{\mathbf{y_{T(n)}}}, \theta\right)|\underline{\mathbf{y_{T(n)}}}\right] = p(\mathbf{y^*_{T(n)+h}}, \ldots, \mathbf{y^*_{T(n)+1}}|\underline{\mathbf{y_{T(n)}}}), \tag{32}$$

by LLN, where $\mathbf{y^*_{T(n)+h}}$ is the true out of sample data point at horizon h. This estimator gives us an idea of how well the model and parameter estimates "fit" the true out-of-sample data. That is, given that $\left\{\mathbf{y^*_{T(n)+h}}, \ldots, \mathbf{y^*_{T(n)+1}}\right\}$ are the actual values observed *ex-post*, what is the probability of their occurrence under our model and *ex-ante* estimated parameters, given the subsample window at iteration n. If one model is more

congruent with the actual future outcomes than another, its *predictive likelihood* should be greater.

When the entire sample data set is known ex-post, we can interpret the sequence of one-step ahead predictive densities as the marginal likelihood:

$$p\left(\underline{\mathbf{y_T}}\right) = \prod_{t=1}^{T^*-1} p\left(\mathbf{y^*_{t+1}}|\underline{\mathbf{y_t}}\right). \tag{33}$$

However, we would like to generalize (33) to account for our iterative sample window schemes discussed above. Instead of assuming that the entire data set is known, rather we will assume that under each sample window, n, only the data $\underline{\mathbf{y}_{\mathbf{T(n)}}}$ is known ex-ante when we estimate the model parameters that define $\hat{p}_n(\cdot)$. Then, keeping in mind the ex-post predictive likelihood as in (32), we can take the product across the N sample windows to obtain a measure of ex-post fit at the first horizon $h = 1$:

$$\prod_{n=1}^{N} \hat{p}_n\left(\mathbf{y^*_{T(n)+1}}|\underline{\mathbf{y_{T(n)}}}\right). \tag{34}$$

Under the recursive window scheme, the expression (34) is the analog to (33), but where we take the product across only a subset of the data, $t = T(1), \ldots, T(N)$, and where we employ the changing estimated predictive distribution $\hat{p}_n(\cdot)$.

However, under the rolling sample window scheme, it is not clear how to interpret (34). Moreover, taking the product across sample windows under changing densities and window schemes forces us to reinterpret (34) not as a density, but purely as a product of forecast metrics. Therefore, taking logs we can transform the product (34) into a sum and interpret this metric as an average across dependent forecast attempts. Moreover, we can also consider robustness of the forecast attempts via sample moments such as

the variance across sample window forecasts. For example, we have the mean of the log-predictive likelihoods as:

$$MLPL_{h=1,N} = \frac{1}{N}\sum_{n=1}^{N} \ln\left\{\hat{p}_n\left(\mathbf{y}^*_{\mathbf{T(n)+1}}|\underline{\mathbf{y}_{\mathbf{T(n)}}}\right)\right\}, \tag{35}$$

where each term in the sum will hereon be denoted as $LPL_{h=1,n}$, and the variance is given as,

$$VLPL_{h=1,N} = \frac{1}{N}\sum_{n=1}^{N}\left(\ln\left\{\hat{p}_n\left(\mathbf{y}^*_{\mathbf{T(n)+1}}|\underline{\mathbf{y}_{\mathbf{T(n)}}}\right)\right\} - MLPL_{h=1,N}\right)^2. \tag{36}$$

Finally, we generalize the above two sample moments to any horizon by replacing the predictive likelihoods with the ex-post one-step ahead predictive distribution at horizon h:

$$\hat{p}_n\left(\mathbf{y}^*_{\mathbf{T(n)+h}}|\mathbf{y}^*_{\mathbf{T(n)+(h-1)}},\ldots,\mathbf{y}^*_{\mathbf{T(n)+1}},\underline{\mathbf{y}_{\mathbf{T(n)}}}\right). \tag{37}$$

That is, the last term in the factorization of the joint predictive likelihood given in (32).

We can compare the $Clark$ and $IWSV$ specifications by looking at how these sample moments differ across forecast horizons. For example, we could compare the difference between the two model's $MLPL_{h,N}$'s across horizons $h = 1,\ldots,H$ as a means of comparing the "term structure" of competing forecast performance across the horizons.

Furthermore, this difference can be "decomposed" into a sum of N log-ratios which can be compared across sample windows $n = 1,\ldots,N$ to suggest at which sample window, n, either model did better, or worse, at forecasting given some fixed horizon h. This difference, given, say, models $A \equiv IWSV$ and $B \equiv Clark$ is defined, for

example given $h = 1$, as

$$N\left(MLPL^{A}_{h=1,N} - MLPL^{B}_{h=1,N}\right) = \sum_{n=1}^{N} \ln\left\{p_{\hat{A},n}\left(\mathbf{y}^{*}_{\mathbf{T(n)+1}}|\underline{\mathbf{y}_{\mathbf{T(n)}}}\right)\right\} - \sum_{n=1}^{N} \ln\left\{p_{\hat{B},n}\left(\mathbf{y}^{*}_{\mathbf{T(n)+1}}|\underline{\mathbf{y}_{\mathbf{T(n)}}}\right)\right\} = \sum_{n=1}^{N} \ln\left\{\frac{p_{\hat{A},n}\left(\mathbf{y}^{*}_{\mathbf{T(n)+1}}|\underline{\mathbf{y}_{\mathbf{T(n)}}}\right)}{p_{\hat{B},n}\left(\mathbf{y}^{*}_{\mathbf{T(n)+1}}|\underline{\mathbf{y}_{\mathbf{T(n)}}}\right)}\right\}. \quad (38)$$

after post multiplication by N. In fact, each log-ratio term in the sum can be interpreted as the ex-post predictive *Bayes factor* in favour of model A over model B, at sample window n.

Bayes factors are the standard Bayesian method of model comparison. The *predictive likelihood* method represents an inherently Bayesian approach to forecast comparison and as such we do not require p-values, since we obtain the finite sample distribution directly. For more details on Bayesian versus frequentist approaches to forecast generation and analysis see Geweke and Amisano (2010).

7 Applications

This section will now discuss applications. The first subsection will evaluate the implementation of the Bayesian estimation methodology via simulated data. The second subsection applies the $Clark$ and $IWSV$ VAR volatility specifications to the macroeconomic data set. Specifically, we will endeavour to compare the two models along a number of dimensions, including: point and interval forecast accuracy, posterior trend estimation, VAR rate of decay, volatility process behaviour, and estimation of the other model parameters.

7.1 Monte-Carlo Analysis

We first perform a Monte-Carlo analysis to provide some insight on the implementation of the Bayesian methodology. We first generate an artificial data set, following a $IWSV$ model. This dataset is used within a rolling sample window scheme to iterate a sequence of estimations of both the $Clark$ and $IWSV$ specifications. The experiment will provide information on the convergence and accuracy of the $IWSV$ based estimation of the parameters. It will also be used to detect the misspecification of Clark's model. Moreover, we will compare forecasts from the misspecified model according to the metrics discussed in Section 6.2.

i) The Data Generating Process

We simulate 370 data points according to the $IWSV$ model. The selected orders are 3 for the VAR component, and 3 for the Inverse Wishart component. The parameter

values are:

$$\mathbf{C} = 0.3\mathbf{I_p}, \tag{39a}$$

$$\mathbf{A_1} = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.75 & 0 & 0 \\ 0 & 0 & 0.85 & 0 \\ 0 & 0 & 0 & 0.98 \end{bmatrix}, \quad \mathbf{A_2} = \mathbf{A_3} = \mathbf{0}, \tag{39b}$$

$$\mathbf{\Pi_1} = \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0 & 0.8 \end{bmatrix}, \quad \mathbf{\Pi_2} = \mathbf{\Pi_3} = \mathbf{0}, \tag{39c}$$

$$\mathbf{\Psi} = \begin{bmatrix} 3.0 & 0 & 2.5 & 0 \end{bmatrix}', \quad \mathbf{d_t} = 1, \forall t, \quad \text{so no trend, and} \tag{39d}$$

$$\nu = 30. \tag{39e}$$

This model implies the following relationships among the series variables, $y_{i,t}$, $i = 1, \ldots, 4$:

$$y_{1,t} = 3.0 + 0.25\left(y_{1,t-1} - 3.0\right) + v_{1,t}, \tag{40a}$$

$$y_{2,t} = 0.8y_{2,t-1} + v_{2,t}, \tag{40b}$$

$$y_{3,t} = 2.5 + 0.8\left(y_{3,t-1} - 2.5\right) + v_{3,t}, \tag{40c}$$

$$\text{and} \quad y_{4,t} = 0.8y_{4,t-1} + v_{4,t}. \tag{40d}$$

The following relationship on the conditional scale matrix, $\mathbf{S_{t-1}}$, of the Inverse Wishart

covariance matrices $\mathbf{\Sigma_t}$, is given as:

$$\mathbf{S_{t-1}} = \begin{bmatrix} 0.09 + 0.25\sigma^2_{11,t-1} & 0.375\sigma_{12,t-1} & 0.425\sigma_{13,t-1} & 0.49\sigma_{14,t-1} \\ 0.375\sigma_{21,t-1} & 0.09 + 0.5625\sigma^2_{22,t-1} & 0.6375\sigma_{23,t-1} & 0.735\sigma_{24,t-1} \\ 0.425\sigma_{31,t-1} & 0.6375\sigma_{32,t-1} & 0.09 + 0.7225\sigma^2_{33,t-1} & 0.833\sigma_{34,t-1} \\ 0.49\sigma_{41,t-1} & 0.735\sigma_{42,t-1} & 0.833\sigma_{43,t-1} & 0.09 + 0.96\sigma^2_{44,t-1} \end{bmatrix} \cdot (30 - 4 - 1), \quad (41)$$

where the scale matrix is given as in equation (7c) above, and the $\sigma_{ij,t-1}$ are the elements of $\mathbf{\Sigma_{t-1}}$. Moreover, from equation (9a), the conditional mean of the stochastic covariance matrix $\mathbf{\Sigma_t}$ is given as $\mathbf{S_{t-1}}/(30 - 4 - 1) = \mathbf{CC}' + \mathbf{A_1 \Sigma_{t-1} A_1}'$. Finally, by applying **Proposition 3.1**, the unconditional means of the stochastic volatilies are given as:

$$\sigma^2_{11} = 0.12, \quad \sigma^2_{22} = 0.21, \quad \sigma^2_{33} = 0.32, \quad \text{and} \quad \sigma^2_{44} = 2.25, \quad (42)$$

where the stochastic covolatilities have zero unconditional mean.

From the proof of **Proposition 3.1** the $IWSV$ model can be written as:

$$\mathbf{\Sigma_t} = \mathbf{CC}' + \mathbf{A_1 \Sigma_{t-1} A_1}' + \mathbf{Z_t}, \quad (43)$$

where $\mathbf{Z_t}$ is a zero mean matrix of weak white noises. Furthermore, this expression can be vectorized and rewritten in terms of the autoregressive coefficient matrix $\mathbf{\Upsilon} = \mathbf{L}\,(\mathbf{A_1} \otimes \mathbf{A_1})\,\mathbf{D}$ as:

$$vech(\mathbf{\Sigma_t}) = vech(\mathbf{CC}') + \mathbf{\Upsilon} vech(\mathbf{\Sigma_{t-1}}) + vech(\mathbf{Z_t}). \quad (44)$$

The persistence in the (co)volatility series are therefore measured by looking at the eigenvalues of the 10×10 dimensional $\boldsymbol{\Upsilon}$ matrix that determines the rate of reversion to the unconditional mean, or response, given a unit impulse shock, $\mathbf{Z}_\tau$, to the (co)volatilies, at time τ. Since, the matrix $\boldsymbol{\Upsilon}$ is diagonal, we can easily solve for the eigenvalues as:

$$\{0.250, 0.375, 0.425, 0.490, 0.563, 0.638, 0.723, 0.735, 0.833, 0.960\}. \qquad (45)$$

The largest eigenvalue is close to a unit root, and so it will influence $\sigma^2_{44,t}$ to exhibit the slowest rate of autoregressive decay.

The following figures describe the simulated data set. **Figure 3**, provides the sample paths of the simulated data series $\mathbf{y_t}$. **Figure 4**, provides a plot of the simulated stochastic volatilities $\sigma^2_{ii,t}$, $i = 1, \ldots, 4$, that is, the diagonal elements of the matrix $\boldsymbol{\Sigma_t}$. Finally, **Figure 5**, provides the i, jth stochastic correlations $\rho_{ij,t} = \frac{\sigma_{ij,t}}{\sqrt{\sigma^2_{ii,t}}\sqrt{\sigma^2_{jj,t}}}$.

These artificial series reveal a number of features. For example, we can see that the values chosen for $\boldsymbol{\Psi}$ influence the mean of the series $\mathbf{y_t}$ in **Figure 3**. Moreover, since the shocks, $v_{i,t}$, are driven by their conditional volatilities, $\sigma^2_{ii,t}$, their conditional mean $E_{t-1}[\sigma^2_{ii,t}] = 0.09 + \alpha_{ii}\sigma^2_{ii,t-1}$ plays a role in determining the magnitude of the volatility of $y_{i,t}$. Indeed, series $y_{i,t}$, associated with stochastic volatilities with larger unconditional means, $0.09/(1 - \alpha_{ii})$, tend to exhibit larger overall episodes of volatility spikes.

For example, from **Figure 3**, we see that the 4th series exhibits the largest volatility episodes as it is associated with $E_{t-1}[\sigma^2_{44,t}] = 0.09 + 0.96\sigma^2_{44,t-1}$, and the largest unconditional mean $\sigma^2_{44} = 2.25$. Moreover, by examining **Figure 3**, we see that the $IWSV$ process exhibits volatility clustering for the 4th series, $y_{4,t}$. For example, volatility be-

tween periods 1 to 100 is much smaller than between periods 200 to 300, and volatility episodes tend to persist across time. Furthermore, these episodes tend to coincide with larger increases in the volatility process $\sigma^2_{44,t}$ given in **Figure 4**, although this relationship is convoluted since it also involves the autoregressive behaviour of the equations given in (40). Finally, the stochastic correlations seen in **Figure 5** tend to grow larger in magnitude as the α^2_{ij} of the corresponding i, jth expected variance components increase, since it represents a multiplicative constant in the conditional variance of the (co)volatility $V_{t-1}[\sigma_{ij,t}]$. For example, the $\rho_{ij,t}$'s associated with the 3rd and 4th series are more variable than those associated with the 1st and 2nd series.

Figure 1: Subsample sequence by recursive window

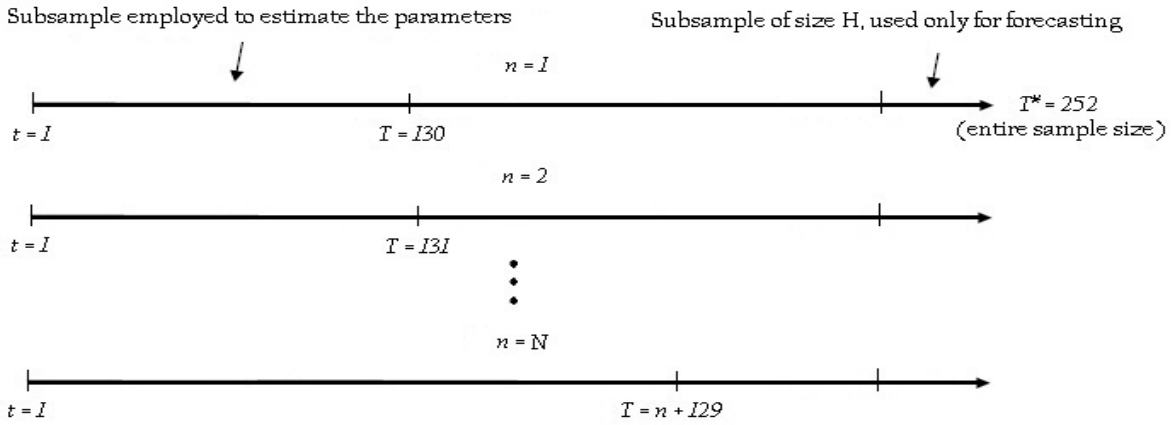


Figure 2: Subsample sequence by rolling window

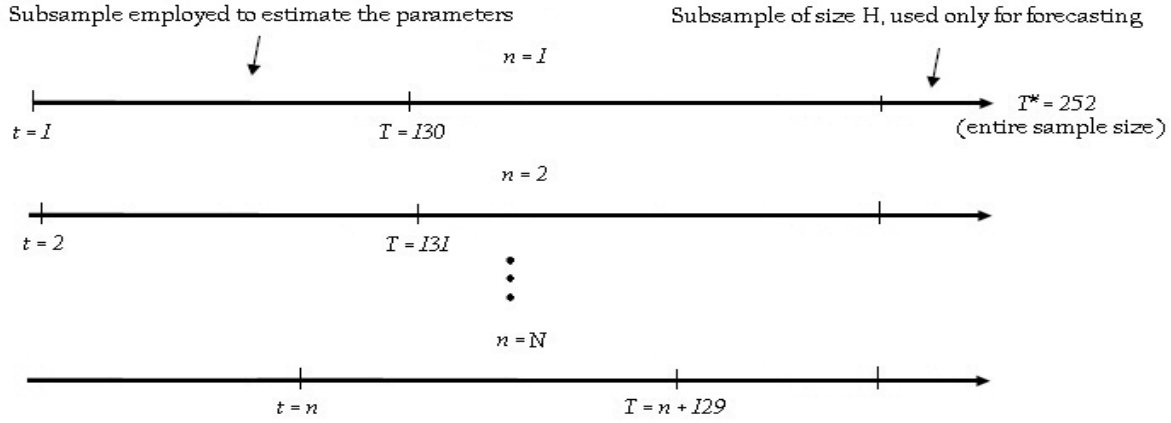


Figure 3: Simulated sample paths $\mathbf{y_t}$

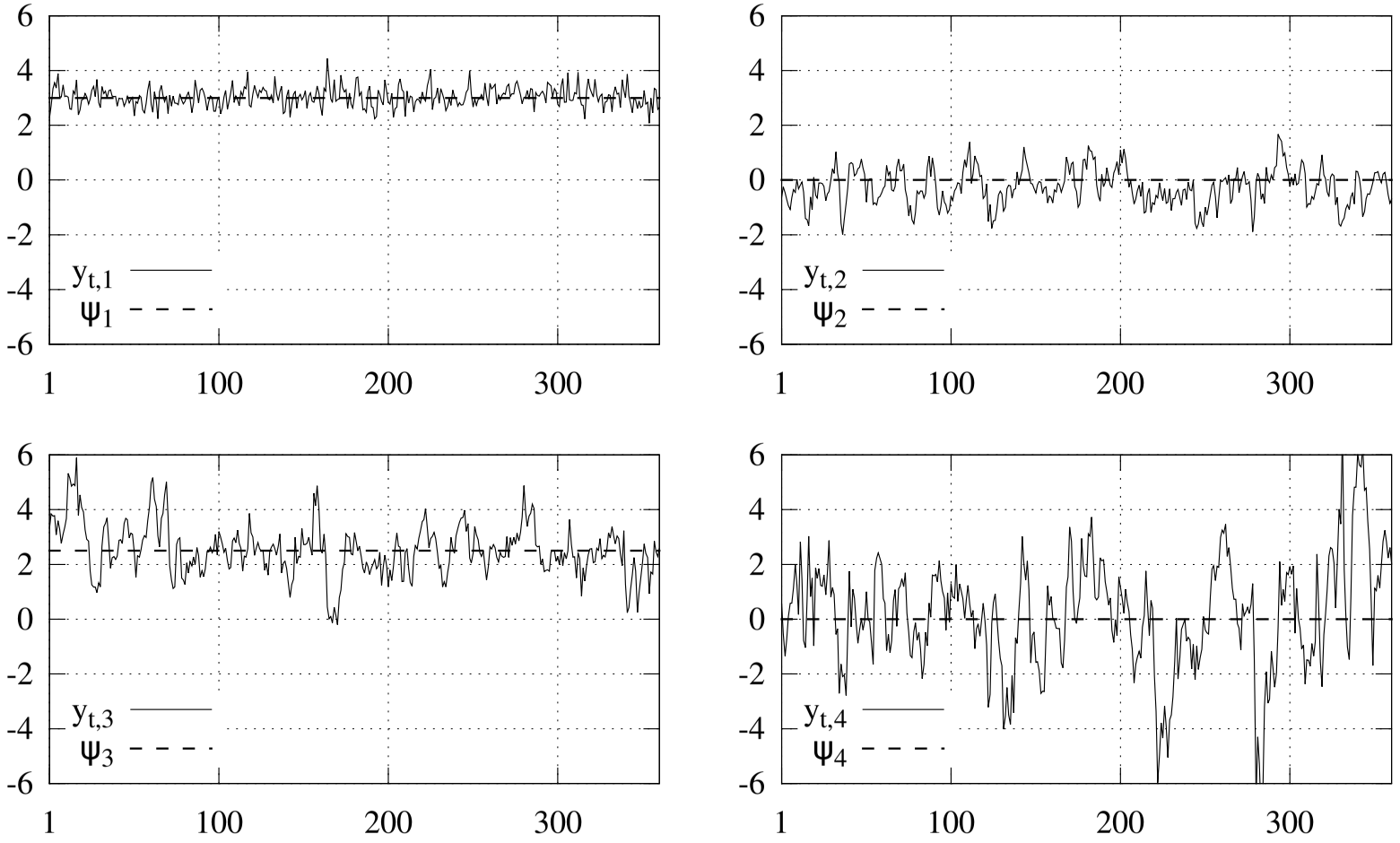


Figure 4: Simulated stochastic volatilities

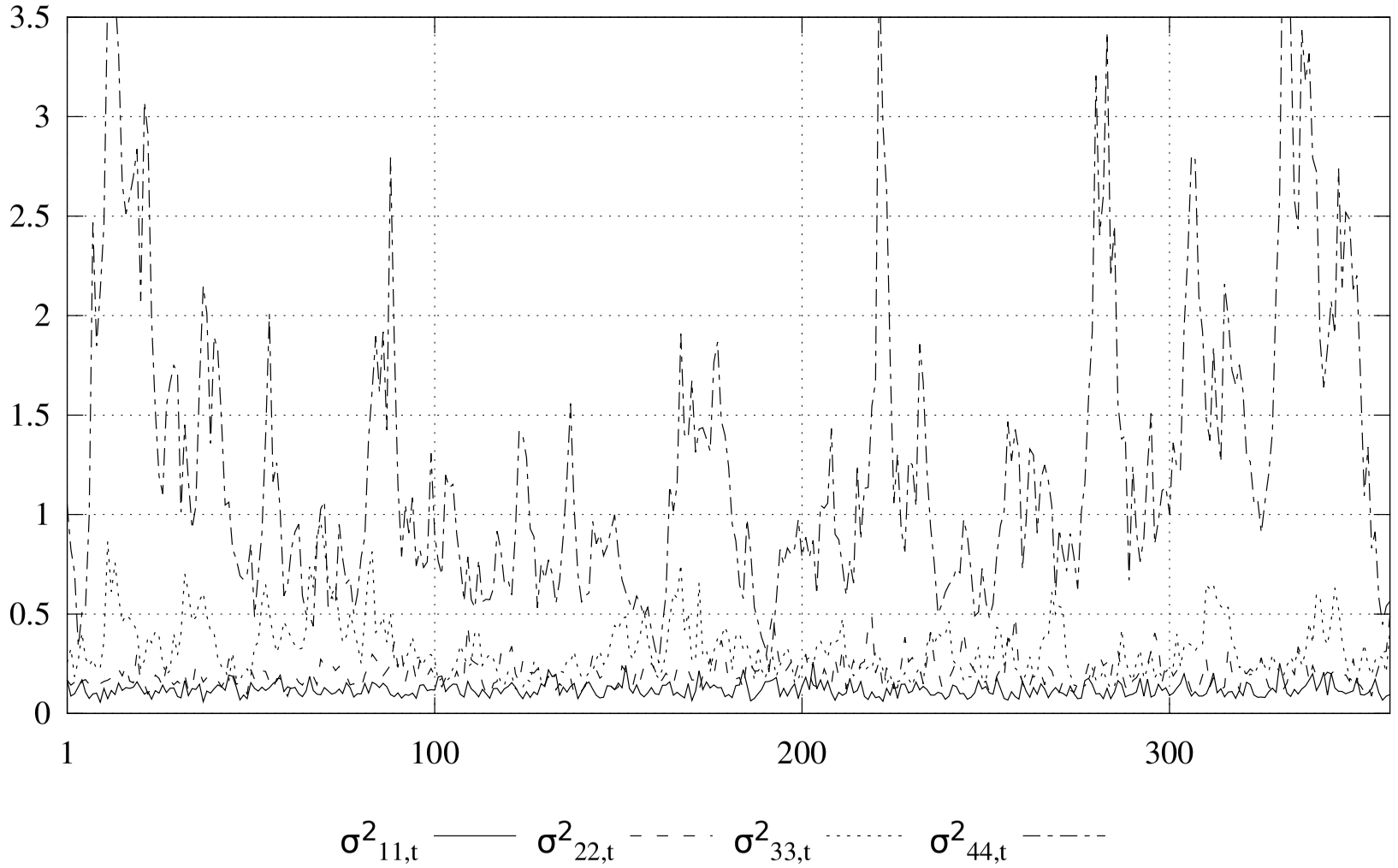
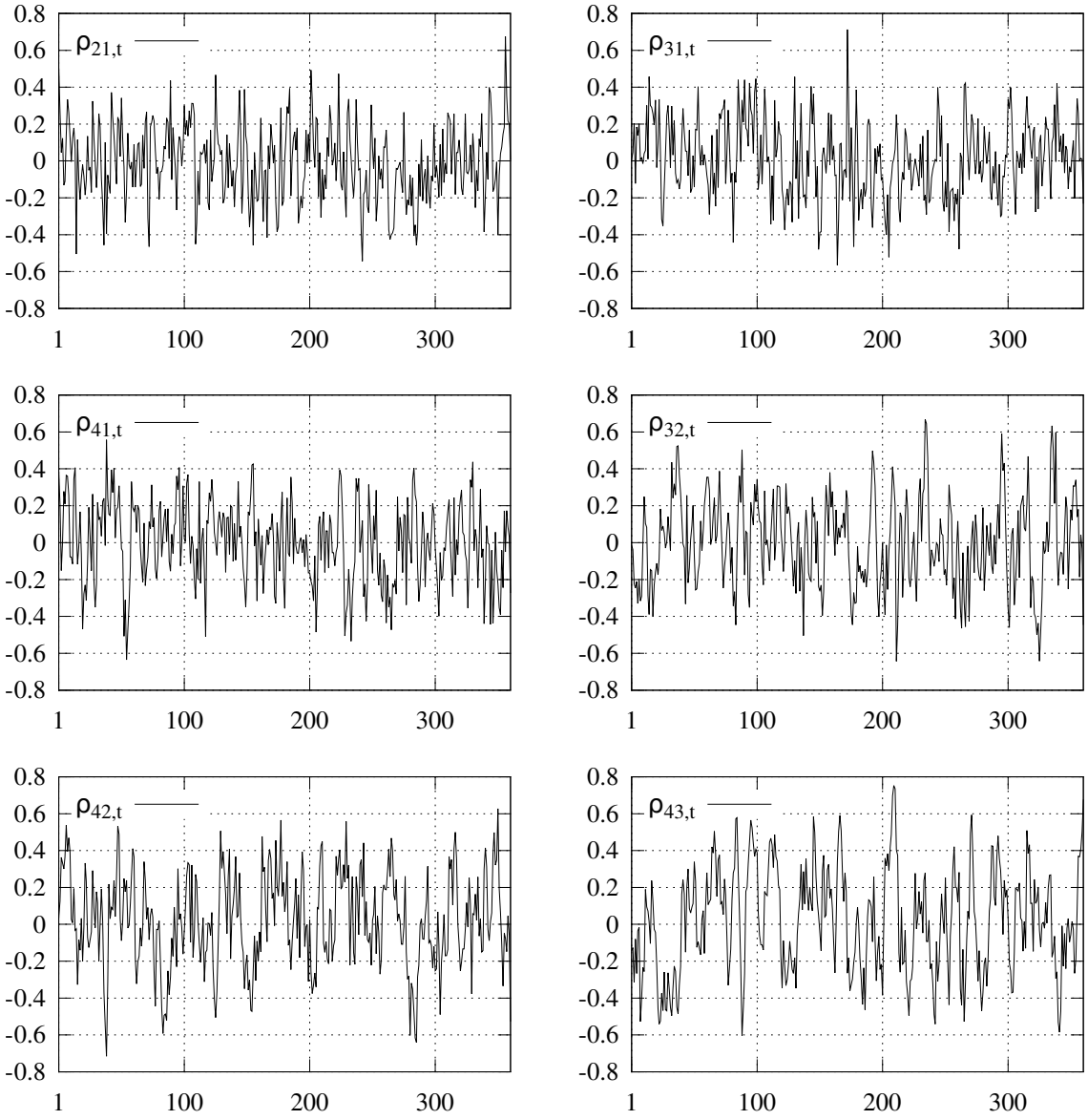


Figure 5: Simulated stochastic correlations



ii) Estimation

Given this simulated data, we estimate both the $Clark$ and $IWSV$ volatility specifications, according to the rolling window scheme, described **Figure 2**, with fixed window

size of 260 and 100 sample windows. The prior densities for the Gibbs sampling are set with means equal to the true values and variances set as in Section 4.

The estimated parameters turn out to be quite close to their true values–see **Tables 3.i** to **4.ii** in Appendix 12, which describe the posterior distribution of the parameters for both models under the 1st sample window, as well as the distribution of the posterior means across all $N = 100$ further iterations. As expected, given the nature of the simulated data, there is less variation in the distribution of the posterior means across sample windows than in the posterior density itself, given the 1st sample window. In other words, there is little change in the posterior means of the model parameters across sample windows.

The only parameter that seems systematically biased is the degree of freedom parameter of the $IWSV$ process, ν. **Figure 6** plots the true value of $\nu = 30$ against the posterior mean and 95% credibility region across the $N = 100$ sample windows. As can be seen, its estimate is lower than the true value. However, as expected, for larger simulated sample sizes the estimated value converges to the true value.

From **Table 4.i**, notice that the posterior means for the VAR parameters of the $Clark$ model are very close to those of the $IWSV$. For the $Clark$ volatility parameters, the true values are omitted since the data were generated using the $IWSV$.

Interestingly, under the $IWSV$ process, the tracking of the latent stochastic volatility estimated posterior means tends to be worse when tracking series associated with the smaller eigenvalues of $\boldsymbol{\Upsilon}$–see **Figures 7.i** and **7.ii** below. It appears as though there exists a lower bound on the tracking of the posterior mean estimates across time for the 1st volatility series—see **Figure 7.i**. Multiple tests have revealed that the smaller the value chosen for the associated diagonal element of $\mathbf{A_1}$, the more pronounced is this lower bound, and conversely, the closer is the diagonal element of $\mathbf{A_1}$ to 1, the better

the posterior mean is able to account for variation in the true volatility sample path. More generally, we assume that the tracking improves as the eigenvalues of the stability matrix $\boldsymbol{\Upsilon} = \sum_{k=1}^{\mathrm{K}} \boldsymbol{\Xi}_{\mathbf{k}}$ approach 1. More investigation is needed, however, to establish definitively the theoretical properties of this phenomenon.

Figure 6: IWSV, Posterior of ν across $N = 100$ sample windows

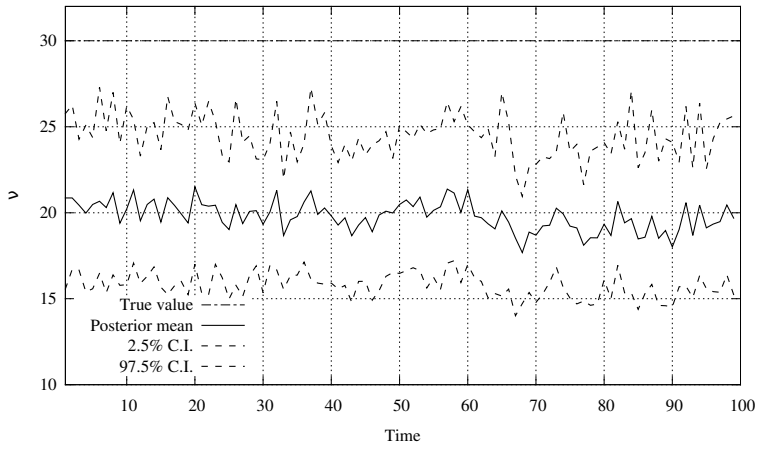
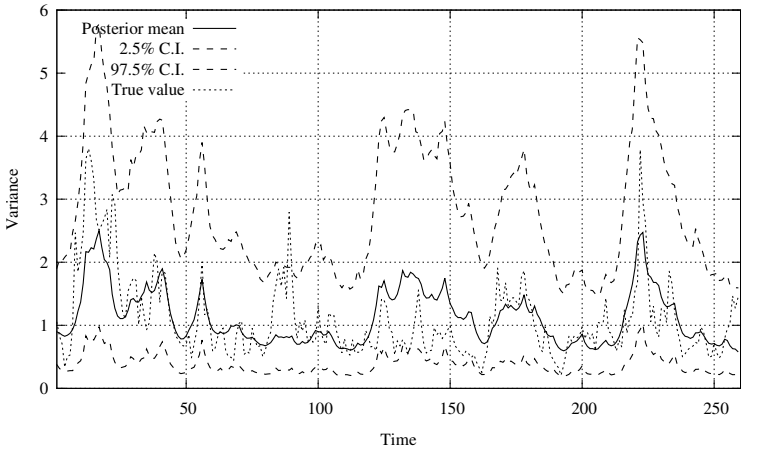


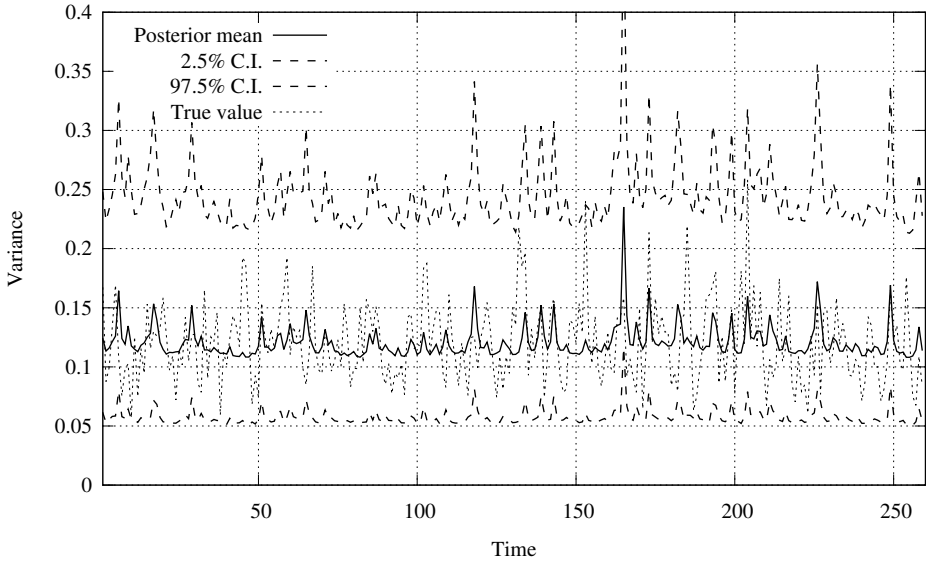
Figure 7.ii: IWSV, filtered latent volatility for 4th series, 1st sample window



iii) Forecasts

In comparing the overall forecast performance, we appeal to the mean of the log predictive likelihoods, $MLPL_{h,N}$, term structure, across horizons $h = 1, \ldots, 10$. Notice

Figure 7.i: IWSV, filtered latent volatility for 1st series, 1st sample window



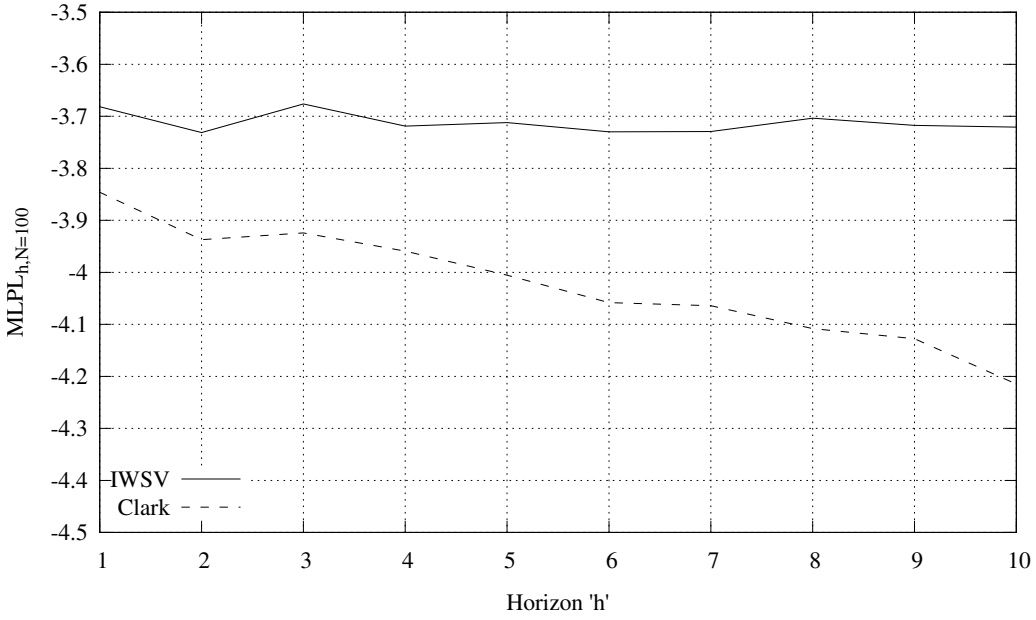
that the $IWSV$ model fits the out-of-sample data much better than the $Clark$, especially at large horizons–see **Figure 8**. At large horizons, such as $h = 10$, the $MLPL_{h,N}$ metric is much larger for the $IWSV$ model, which suggests that this model fits the out-of-sample data better as the horizon increases.

Of course, we can also consider the term structure of the model specific log predictive likelihoods, $LPL_{h,n}$, across the individual sample windows. In this case we can interpret the log-ratios at each sample window, n, as predictive Bayes factors (recall equation (38) from Section 6.2.iii). **Figure 9** plots these Bayes factors across sample windows where model $A \equiv IWSV$ and model $B \equiv Clark$. Larger values suggest that the $IWSV$ is more representative of the out-of-sample outcomes than the $Clark$ at each sample window $n = 1, \ldots, N$.

Figure 9 also suggests that there is a variability in the forecasting performance according to the $LPL_{h,n}$ metric. This Figure 9 provides the term structure of model performance, according to the $LPL_{h,n}$ metric, across the sample windows $n = 1, \ldots, N$,

where as usual $A \equiv IWSV$ and $B \equiv Clark$. The variability in the $LPL_{h,n}$ forecast metrics for each model, can be measured by considering their sample moments across sample windows, given in **Table 1**. Interestingly, while the $IWSV$ fares better in terms of the sample mean of $LPL_{h,n}$ metrics across sample windows, the $Clark$ model metrics have the advantage of being less variable, skewed, and leptokurtic. **Figure 10** provides histograms of these distributions across sample windows, at the 10th horizon.

Figure 8: Simulated data, Forecast horizon term structure according to $MLPL_{h,N}$ metric, $N = 100$ sample windows



7.2 Real data

i) The Clark (2011) data set

Let us now turn to the real-world Clark (2011) macroeconomic data set. In this case we do not know the true model and so we will attempt to choose between the two

Figure 9: Simulated data, Sample window term structure according to difference of $LPL_{h,n}$'s metric, $h = 10$

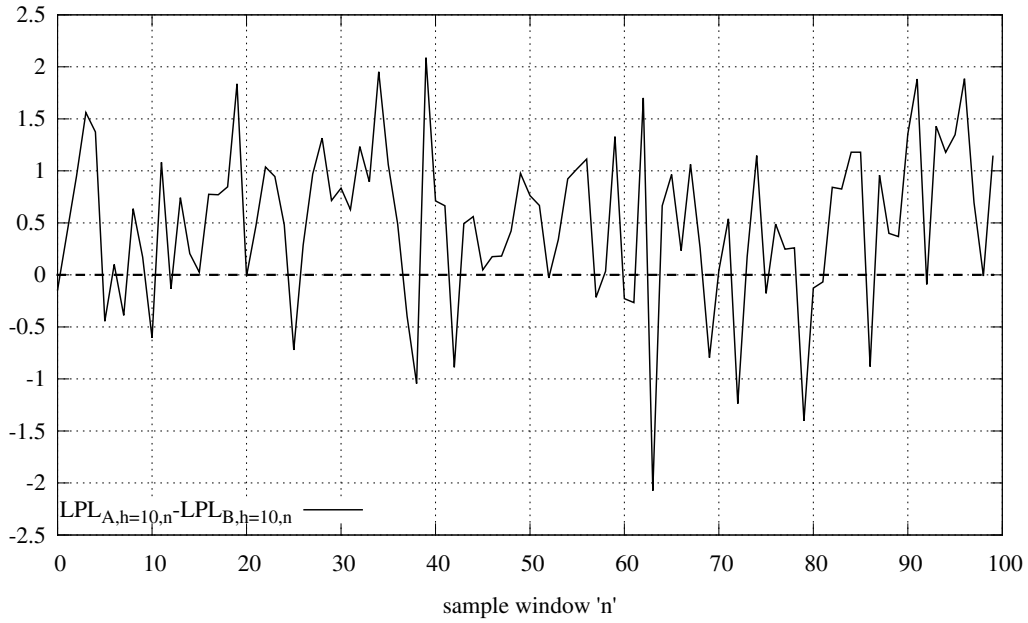
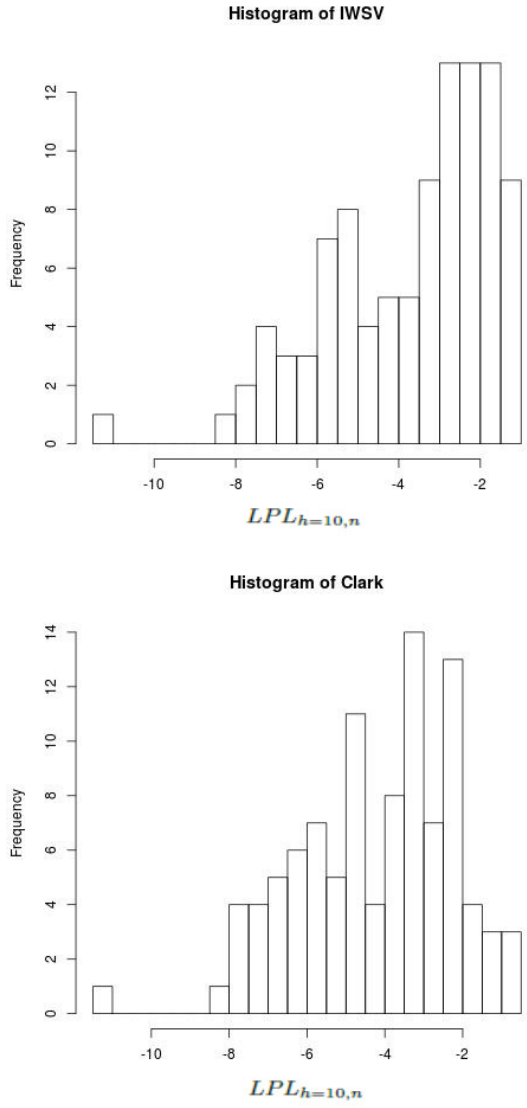


Table 1: Simulated data, Sample moments of the $LPL_{h,n}$ metrics across $N = 100$ sample windows

	Horizon $h = 1$		Horizon $h = 10$	
	IWSV	Clark	IWSV	Clark
mean	-3.6816	-3.8461	-3.7208	-4.2147
stnd. dev.	2.1310	1.9748	2.0290	1.9748
skewness	-1.005	-0.6926	-0.9314	-0.5453
kurtosis	3.834	3.1393	3.5944	3.1594

Figure 10: Histograms of the $LPL_{h=10,n}$ metrics across $n = 1, \ldots, N$ sample windows, 10th horizon



misspecified models, i.e. the $Clark$ and $IWSV$ models, according to out-of-sample forecast performance.

We first consider plots of the Clark (2011) data series provided in figures **Figures 11** and **12**. **Figure 11** provides a plot of the raw data series, along with exponentially smoothed trends, as described in Section 2 above. **Figure 12** presents the data series after detrending by the associated smoothed trend. Recall that the trends applied are those from Clark (2011) in order to replicate the results from that paper, and not necessarily because we believe these trends to provide the best fit.

ii) Estimation

We estimate both volatility models on an initial subsample data size of 130, across 100 sample windows. Both rolling and recursive window schemes are estimated. Various orders of both the VAR and $IWSV$ specification were tested and three lags were ultimately chosen as a balance between parameterization and improvement in model fit. The Gibbs sampling steps in Appendix 11 are performed for both models with a draw size of $M = 100,000$, and a burn-in of $M_0 = 10,000$, and the priors are set as described in Section 4.

Summary statistics of the posterior distributions of the parameters are given in **Tables 5.i** to **6.iii**, in Appendix 12. **Tables 5.i** and **6.i** provide the posterior means and 95% credibility intervals for both the $IWSV$ and $Clark$ model parameters, respectively, given the 1st sample window of a recursive window scheme. **Tables 5.ii** and **6.ii** provide the sample means and 95% confidence intervals for the distribution of the posterior means of both the $IWSV$ and $Clark$ model parameters, across the $N = 100$ sample windows, again given a recursive window scheme. Finally, **Tables 5.iii** and **6.iii** are the analogs to the previous described tables, except we instead employ a rolling sampling window scheme.

These tables reveal a number of interesting features. First, irrespective of sample window size or scheme, the posterior means of many parameters deviate from our assumptions on the prior means, suggesting that the data are informative. For example, within the context of the $IWSV$ model, the elements of the main diagonal of the $\mathbf{C}$ matrix deviate from the assumption of 0.3, although the off-diagonal elements tend to stay close to zero.

Figure 11: Clark (2011) macroeconomic dataset, series and smoothed trends

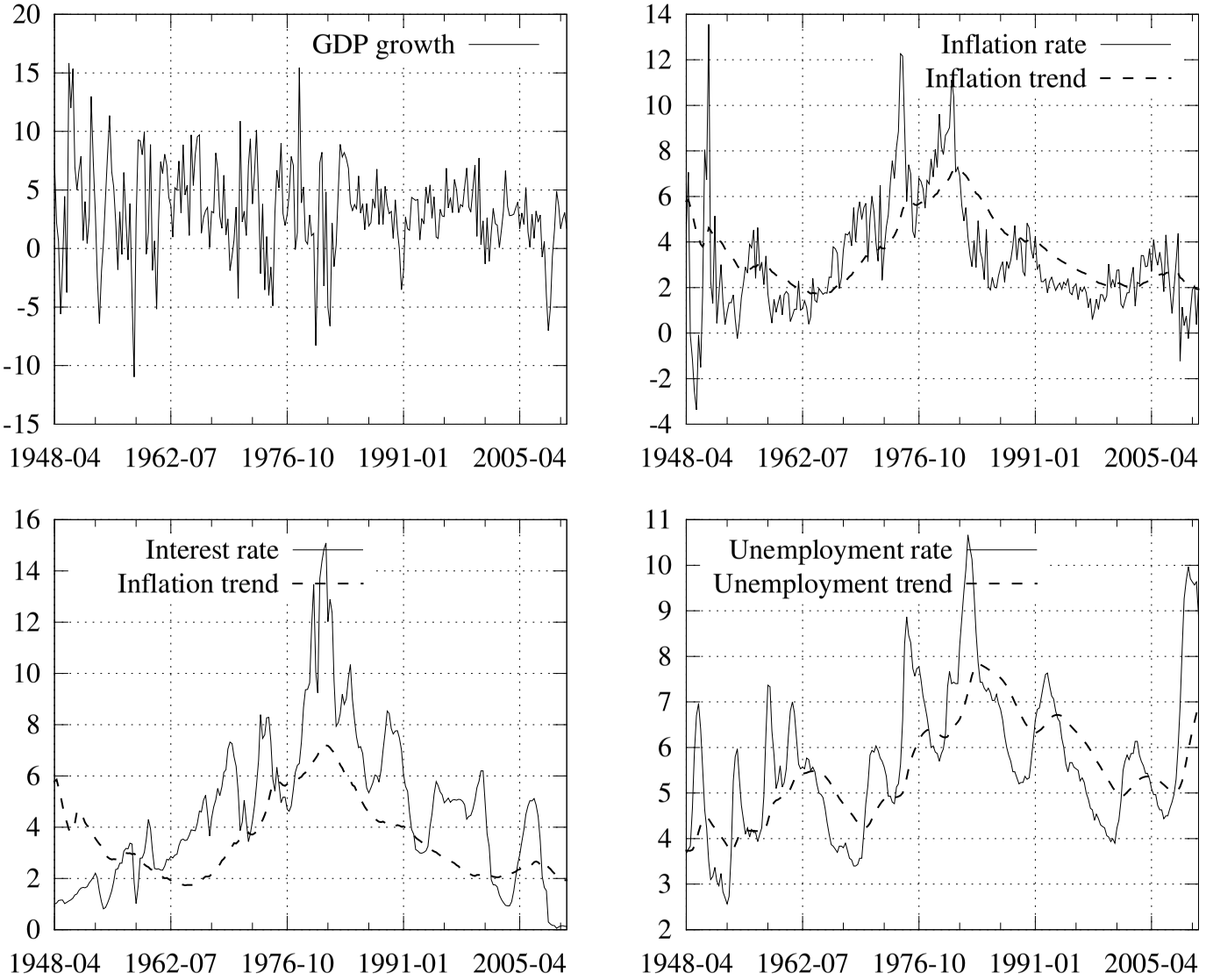
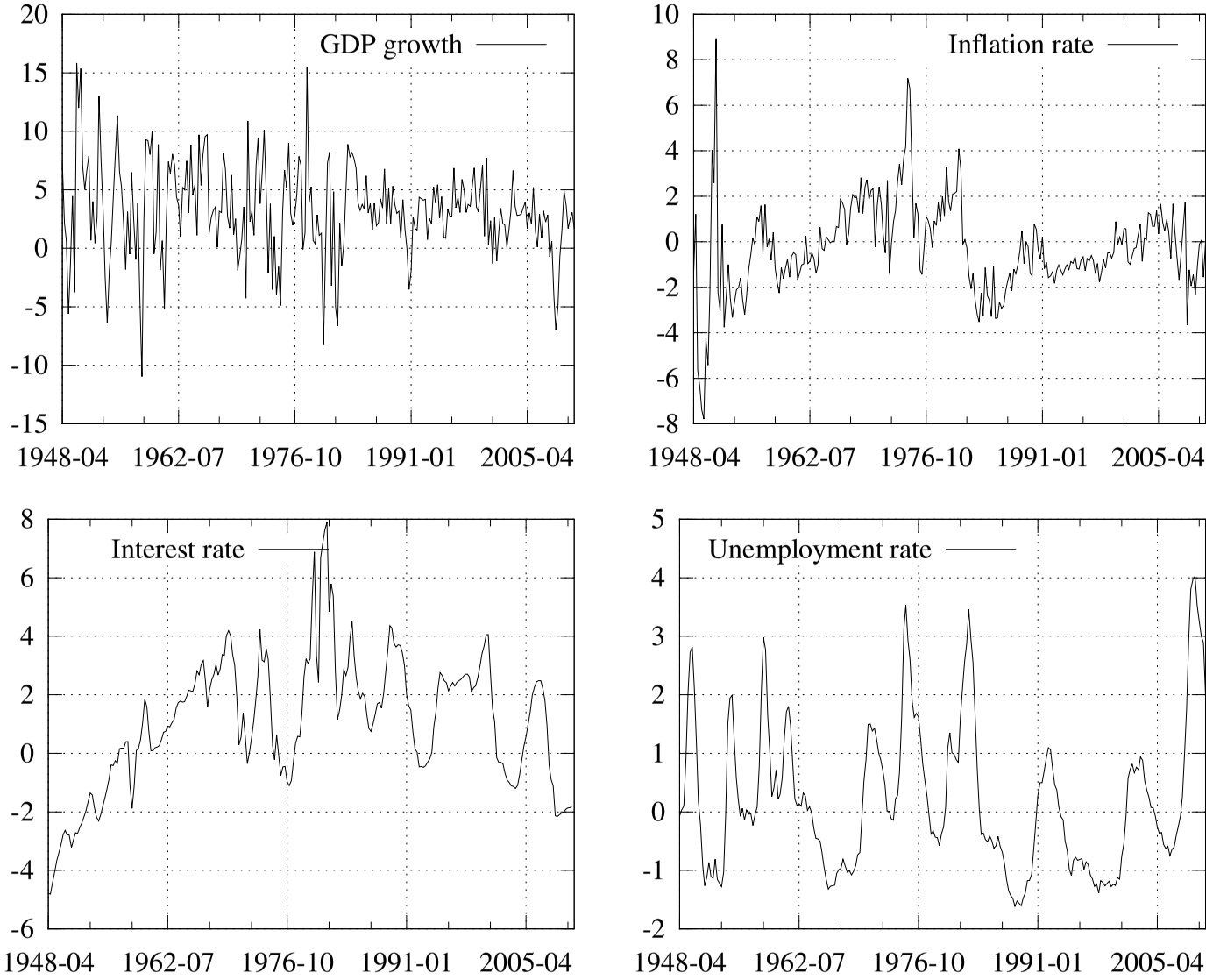


Figure 12: Clark (2011) macroeconomic dataset, detrended series



Interestingly, the degree of freedom parameter ν exhibits posterior means in a range between approximately 15 to 19, higher than our assumption of 15 on the prior. As for the $Clark$ model, the distribution of the posterior means for the elements of the $\mathbf{B}$ matrix, suggests that we can reject the prior mean assumption that these elements are zero. Within the context of both models, the 1st element, ψ_1, of $\mathbf{\Psi}$, associated with GDP growth, exhibits posterior means slightly above the assumption of 3 and the 3rd value, ψ_3, associated with the interest rate, exhibits posterior means slightly below the assumption of 2.5. Moreover, ψ_2 and ψ_4 both exhibit the possibility of non-zero means, despite our assumptions.

Moreover, there is a surprising level of consistency in the evolution of the posterior means across sample windows and window types. For example, let us consider **Figures 23.i** to **24.ii**, in Appendix 12, which plot the posterior distributions of various model parameters from both the $IWSV$ and $Clark$ volatility specifications, across the sample windows for both the recursive and rolling window schemes. The posterior distributions involving the rolling sample windows are much more variable than those that employ the recursive sample window scheme.

We can also check the stability properties of both the VAR and $IWSV$ volatility process across the sample windows. **Figures 13.i** and **13.ii** plot the absolute value of the largest eigenvalues from both the VAR(3) companion matrix given as

$$\begin{bmatrix} \mathbf{\Pi_1} & \mathbf{\Pi_2} & \mathbf{\Pi_3} \\ \mathbf{I_4} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I_4} & \mathbf{0} \end{bmatrix} \tag{46}$$

and the $\mathbf{\Upsilon} = \sum_{k=1}^{3} \mathbf{L}(\mathbf{A_k} \otimes \mathbf{A_k})\mathbf{D}$ matrix, which determines the $IWSV$ stability (re-

call equation (44) and Proposition 3.1), across the $N = 100$ runs. We employ the posterior mean of the relevant parameters in constructing these matrices and their associated eigenvalues. The VAR(3) processes are generally stable. However, it appears that there may exist a unit root in the $IWSV$ volatility process.

Finally, **Figure 14** plots the posterior mean of the latent stochastic volatilities for both the $IWSV$ and $Clark$ models, for the complete sample, given augmented parameters filtered given a recursive sample window at the $N = 100$th iteration. Moreover, **Figure 15.i** and **Figure 15.ii** plot the associated stochastic correlations for both the $IWSV$ and $Clark$ models respectively. The $Clark$ model exhibits stochastic correlations which are too smooth, and, as expected, both models exhibit a negative stochastic correlation, $\rho_{41,t}$, between shocks to GDP growth and the unemployment rate.

Figure 13.i: Largest eigenvalue of VAR(3) companion matrix, across $N = 100$ sample windows

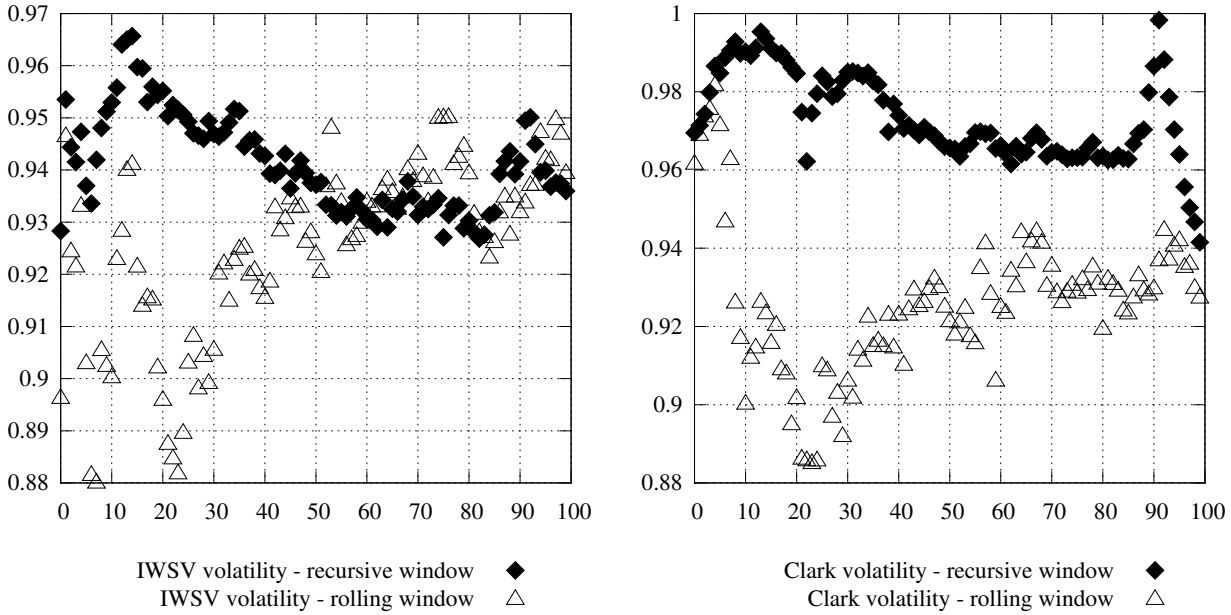


Figure 13.ii: Largest eigenvalue of the Υ matrix, across $N = 100$ sample windows

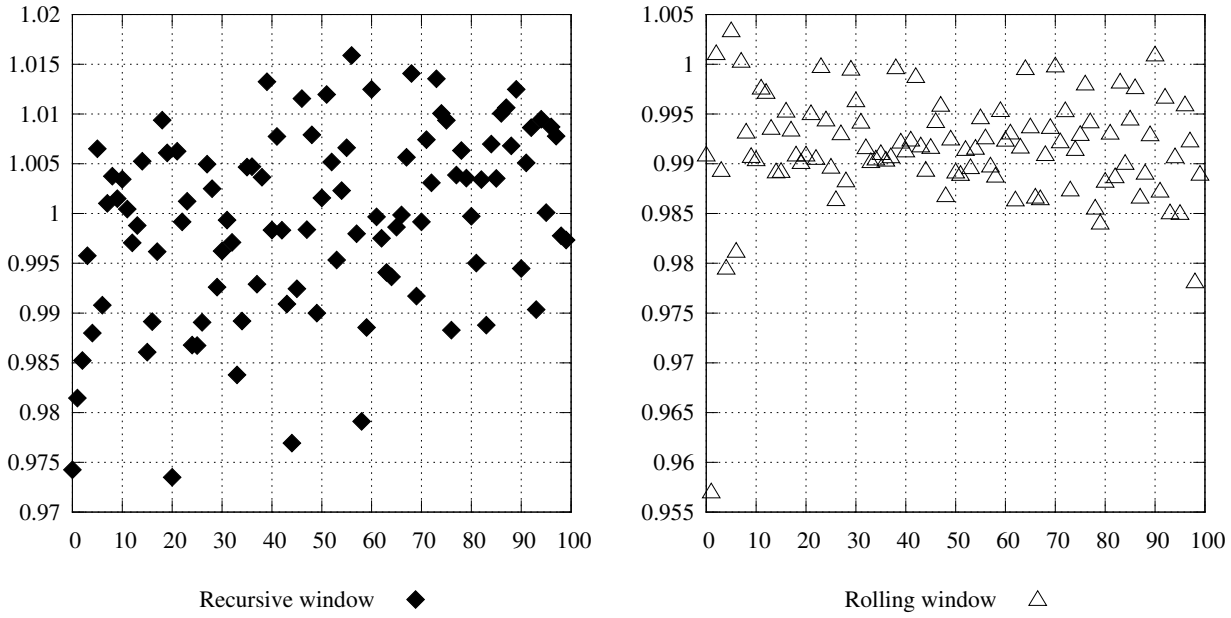
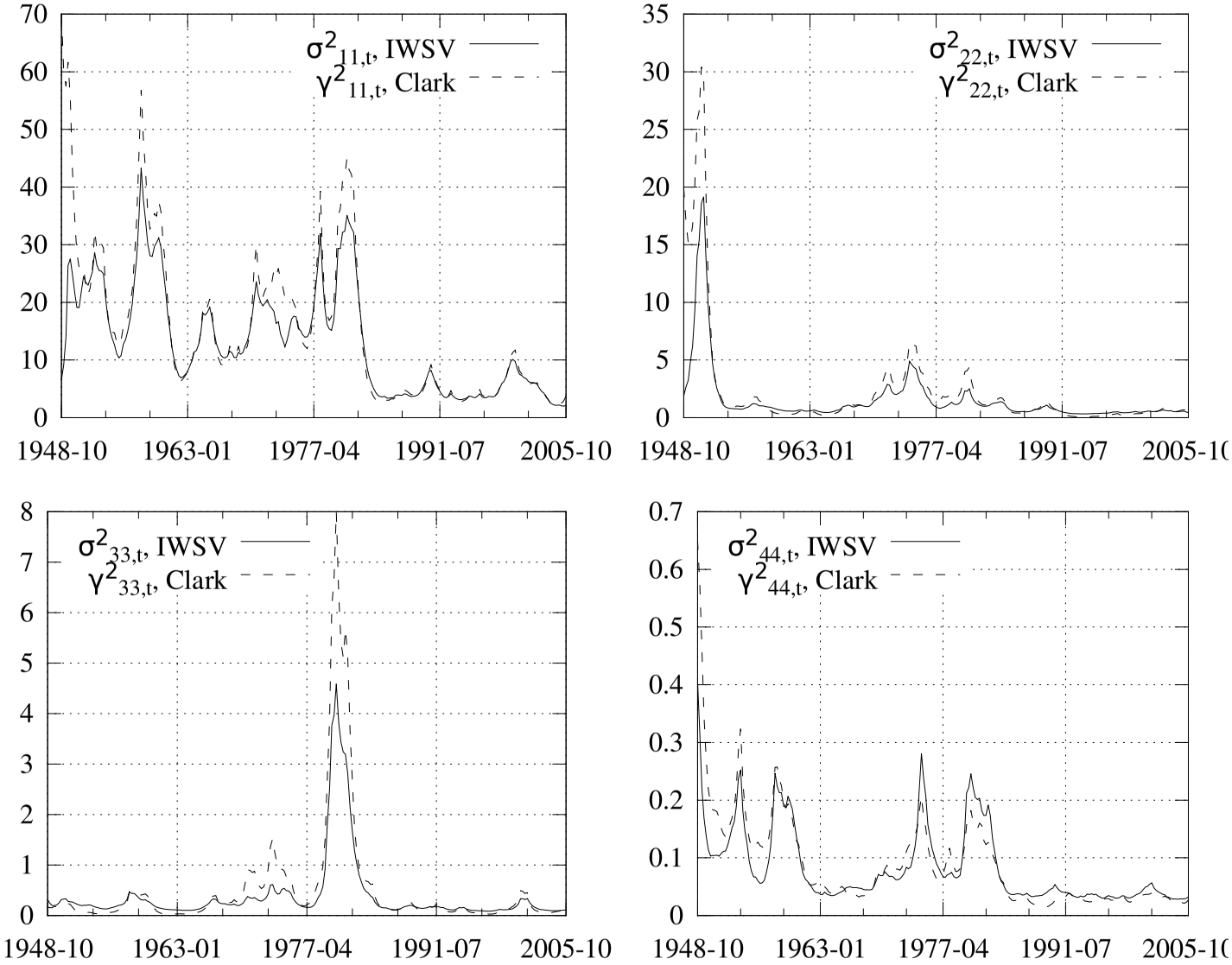


Figure 14: IWSV and Clark, filtered latent stochastic volatilies, 100th sample window, Recursive window



iii) Forecasts

We would now like to establish the forecast properties of the VAR associated with both the $IWSV$ and $Clark$ volatility models. We first compare the $MSE_{h,N}$ metrics which establish point forecast accuracy across the term structure of forecast horizons, $h = 1, \ldots, 20$. **Figure 16** provides plots of the percentage difference in the main diagonal elements of the $MSE_{h,N}$ matrix, across h, for both the recursive sample window scheme and the rolling sample window scheme, respectively.

From these plots it is not clear that either model performs better in terms of point forecast performance. This is to be expected of course, since our $IWSV$ modification to the model is an alternative specification on the volatility of the process, not the mean.

Figure 15.i: Real data, IWSV model, filtered latent stochastic correlations for the complete sample, $n = 100$

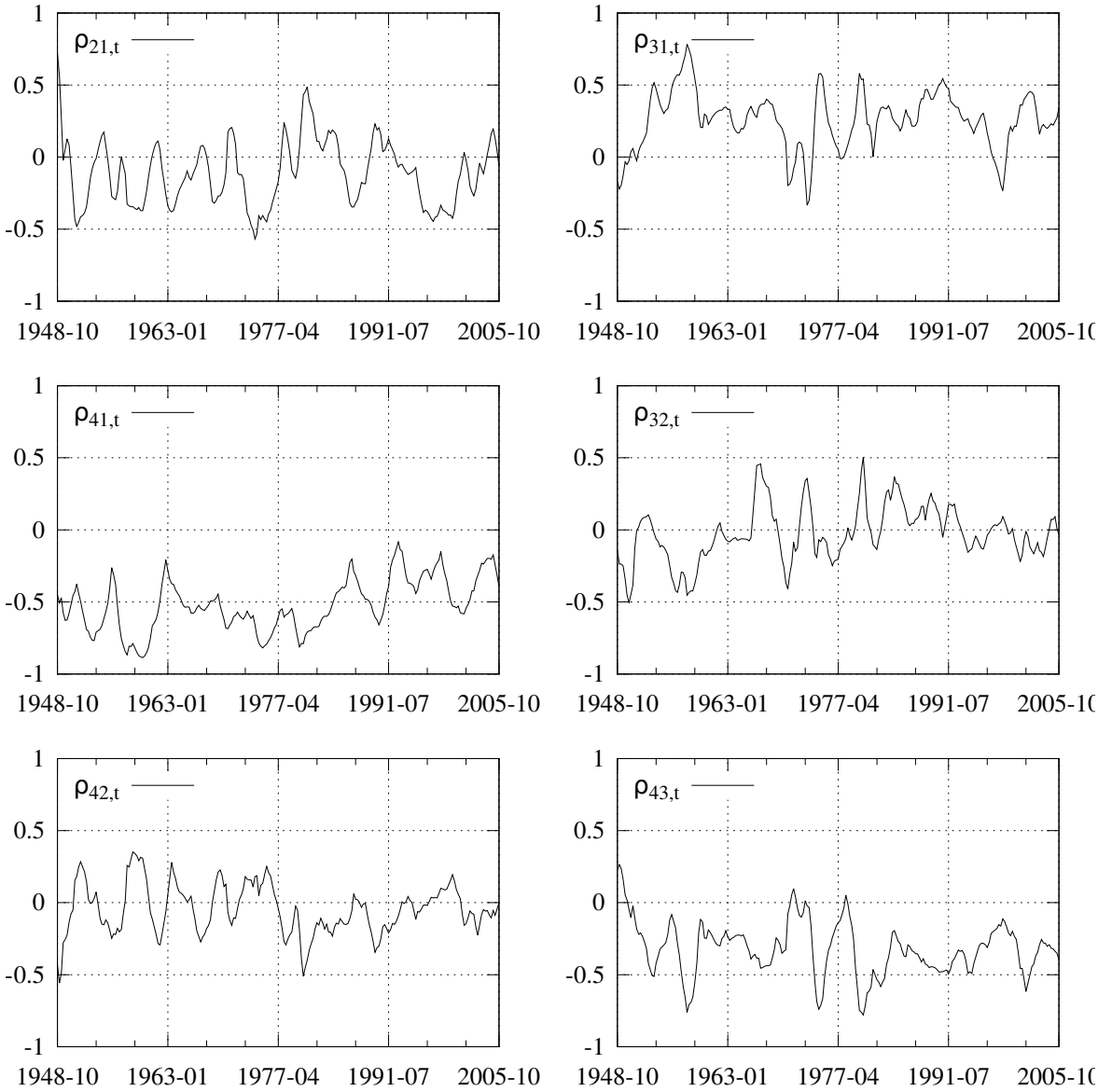


Figure 15.ii: Real data, Clark model, filtered latent stochastic correlations for the complete sample, $n = 100$

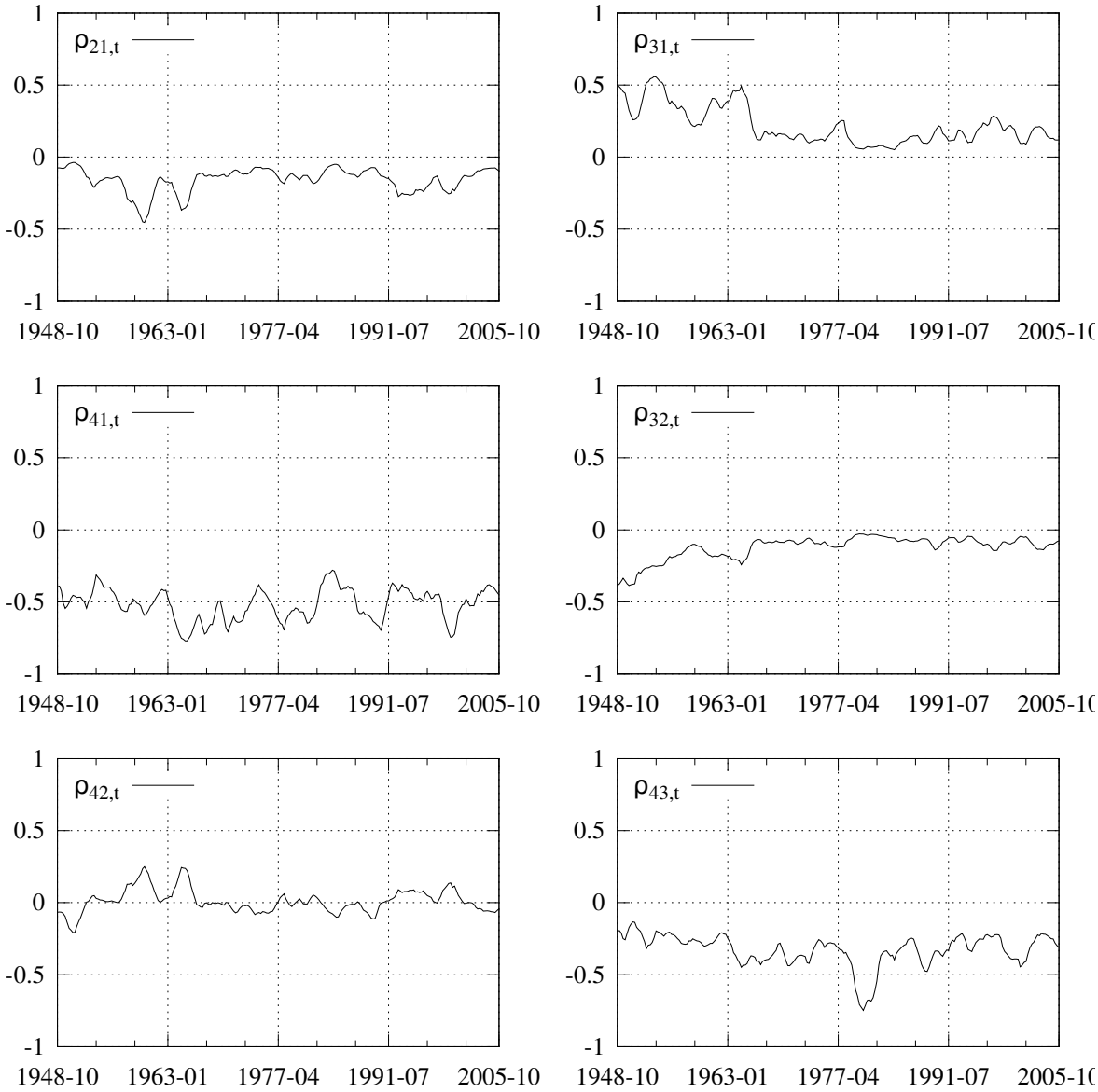
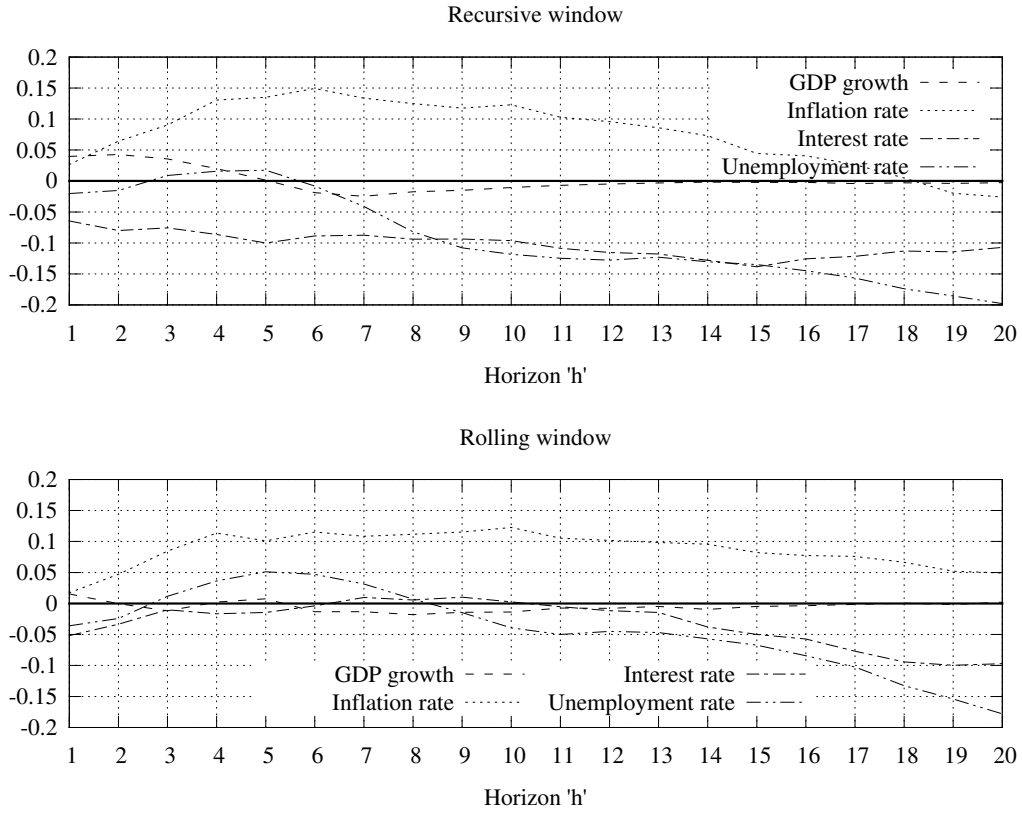


Figure 16: Real data, MSE comparison of VAR forecasts, % difference, both window types (below 0, $IWSV$ better)



Turning now to comparing overall forecast performance, we again appeal to the mean of the log predictive likelihoods, $MLPL_{h,N}$, term structure, across horizons $h = 1, \ldots, 20$. This term structure should illustrate improved out-of-sample predictive likelihood, given the alternative specification on the second moment of the VAR process shocks.

From **Figure 17.i** we see how the mean of these $LPL_{h,n}$ measures is improved as the forecast horizon extends out to H. This result is irrespective of sample window scheme, although the recursive window, which grows in size across the iterations, tends to fare slightly better than the rolling window at further horizons. Interestingly, the

Figure 17.i: Real data, Forecast horizon term structure according to $MLPL_{h,N}$ metric, $N = 100$ sample windows, both Recursive and Rolling sample windows

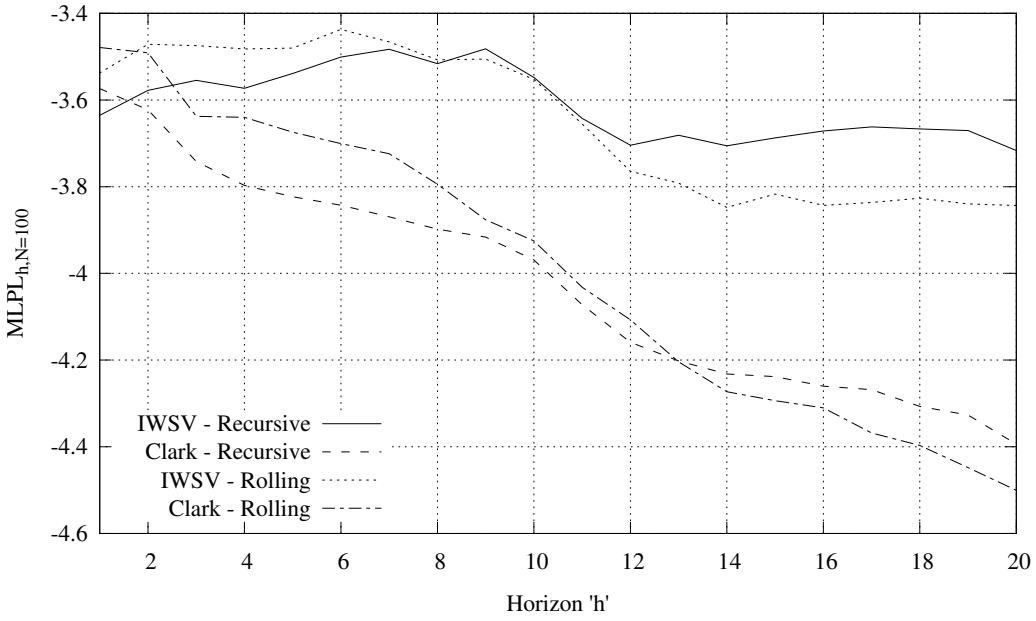
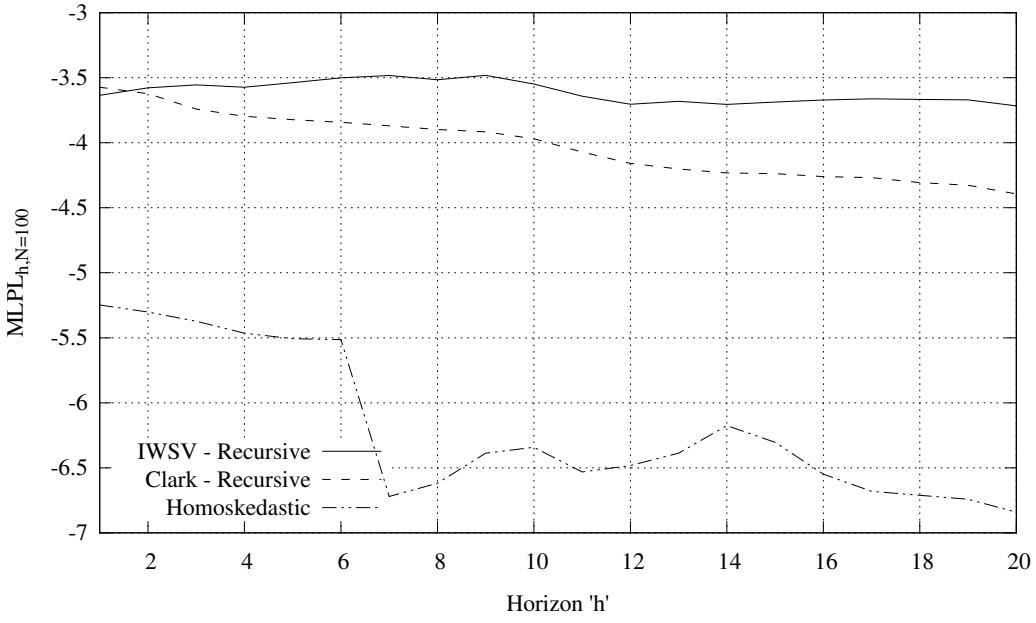


Figure 17.ii: Real data, Forecast horizon term structure according to $MLPL_{h,N}$ metric, $N = 100$ sample windows, includes homoskedastic $\mathbf{v_t}$



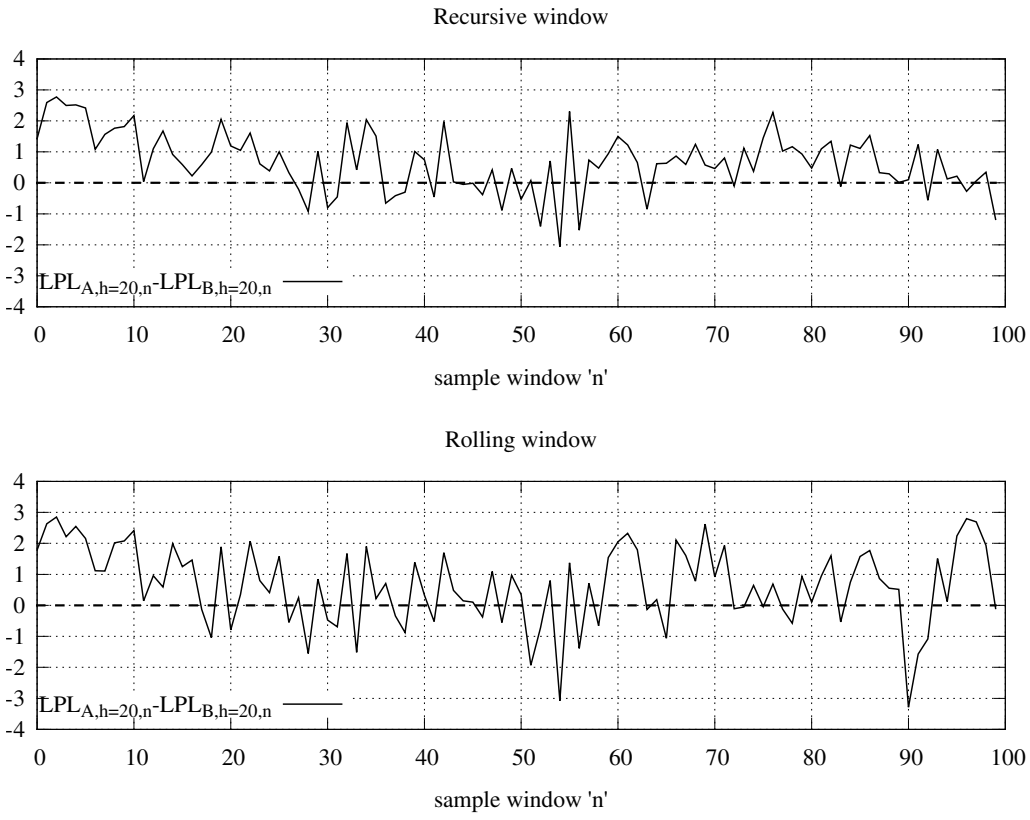
$Clark$ model does better according to this metric at the 1st horizon, across both sample window types. It is not clear why this is the case. For reference, **Figure 17.ii** duplicates the previous figure, but includes the case of homoskedastic VAR innovations, $\mathbf{v_t}$, for reference.[18]

Again, we can also consider the term structure of the model specific log predictive likelihoods, $LPL_{h,n}$, across the individual sample windows. **Figure 18** plots the difference in the $LPL_{h=20,n}$ metrics at horizon, $h = 20$, across sample windows, where model $A \equiv IWSV$ and model $B \equiv Clark$. Larger values suggest that the $IWSV$ is more representative of the out-of-sample outcomes than the $Clark$ at each sample window $n = 1, \ldots, N$.

Moreover, these $LPL_{h,n}$ metrics are changing across the sample windows according to their own sampling distribution, given in **Table 2**. Again, we see similar results as in the simulated data case above in Section 7.1.iii. For the 20th forecast horizon, $h = H = 20$, we find that the mean of the log predictive likelihoods, $MLPL_{h,N=100}$, are larger under the $IWSV$, which initially suggests that the $IWSV$ improves out of sample fit at the longer horizons. However, for the 1st horizon, we see the opposite, that is, the $Clark$ has a larger $MLPL_{h=1,N=100}$ metric associated with it. As for the the other sample moments, generally they suggest larger deviations in forecast performance across the sample windows when using the $IWSV$ model. The larger kurtosis and negative skew imply that the $IWSV$ is more sensitive to rare occurrences of poor forecasting fit, this sensitivity increases as the horizon extends out, and it is worse under

[18]Estimating the VAR model with homoskedastic shocks $\mathbf{v_t}$ is accomplished by replacing the Gibbs sampler steps for the volatility parameters with a single Gibbs step that employs an Inverse Wishart prior density. Since the Inverse Wishart prior is conditionally conjugate with the multivariate Normal, the conditional posterior is also Inverse Wishart. That is, if $\pi(\mathbf{\Sigma}) \sim IW(a_0, \mathbf{V}_0) \Rightarrow p(\mathbf{\Sigma} \mid \mathbf{v}) \sim IW(a_1, \mathbf{V}_1)$ such that $\mathbf{V}_1 = \sum_{t=1}^{T} \mathbf{v_t}\mathbf{v_t}' + \mathbf{V}_0$ and $a_1 = T + a_0$, where $\mathbf{v_t}$ is the $p \times 1$ vector of VAR residuals. $\mathbf{V}_0$ is set to the unconditional sample covariance matrix of simulated VAR residuals (generated with reasonable guesses on the VAR parameters) and $a_0 = 15$.

Figure 18: Real data, Sample window structure of the difference of $LPL_{h,n}$'s metric, $h = 20$



the rolling sample window scheme. Interestingly, with the exception of the recursive sample window scheme at horizon, $h = 1$, the $Clark$ model now exhibits a larger standard deviation of $LPL_{h,n}$, metrics, than the $IWSV$. Finally, to get an idea of the shape of these distributions, **Figure 19** provides histograms of the $LPL_{h,n}$, metrics across the sample windows.

Figure 20 provides the analog to **Figure 18**, in scatter plot format. That is, it presents the $LPL_{h=20,n}$ values, for each sample window $n = 1, \ldots, 100$, at the largest horizon, $h = 20$. The shape of the scatter gives us an intuition on how each model performs across the sample windows which complements the previous Figure 20 which presented the sample window $LPL_{h=20,n}$ differences in chronological order.

Finally, **Figures 21.i** to **21.iv** plot the out-of-sample forecasts of the VAR series, $\mathbf{y_t}$, out $H = 20$ periods, given the last recursive sample window $n = 100$ (this subsample includes nearly the entire data set). While the forecasted conditional means are quite similar between the $IWSV$ and $Clark$ models, as expected, the prediction intervals are distinctly shaped, reflecting the different underlying stochastic volatility processes. The $Clark$ prediction intervals tend to "bell" out and expand as the horizon grows large, while the $IWSV$ prediction intervals tend to stabilize. Interestingly, in referencing **Figure 18** for this particular sample window, $n = 100$, we see that this represented a subsample where the $Clark$ model performed better. Clearly, in this case the 95% prediction intervals encompass the true data outcome more accurately than the $IWSV$ prediction intervals.

Figure 19: Histograms of the $LPL_{h=20,n}$ metrics across $n = 1, \ldots, N$ sample windows, 20th horizon

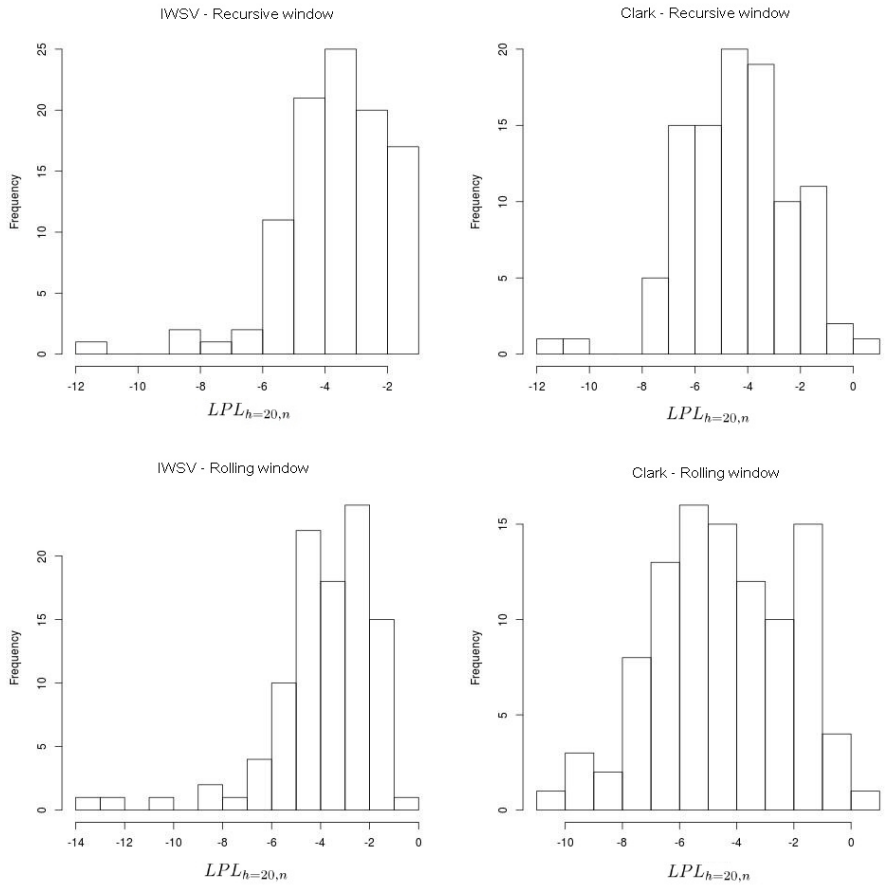


Table 2: Real data, Sample moments of the $LPL_{h,n}$ metrics across $N = 100$ sample windows

Recursive sample window	Horizon $h = 1$		Horizon $h = 20$	
	IWSV	**Clark**	**IWSV**	**Clark**
mean	-3.6352	-3.5736	-3.7165	-4.3938
stnd. dev.	2.0322	1.9614	1.7725	2.0413
skewness	-0.9329	-0.2877	-1.3387	-0.4069
kurtosis	4.1266	2.401	6.5644	3.9734
Rolling sample window	Horizon $h = 1$		Horizon $h = 20$	
	IWSV	**Clark**	**IWSV**	**Clark**
mean	-3.5373	-3.4788	-3.8435	-4.501
stnd. dev.	2.0205	2.0533	2.1202	2.3618
skewness	-0.9310	-0.2463	-1.8423	-0.2015
kurtosis	4.1083	2.1766	8.0861	2.5614

Figure 20: Real data, Sample window structure of the $LPL_{h,n}$'s metric, $h = 20$

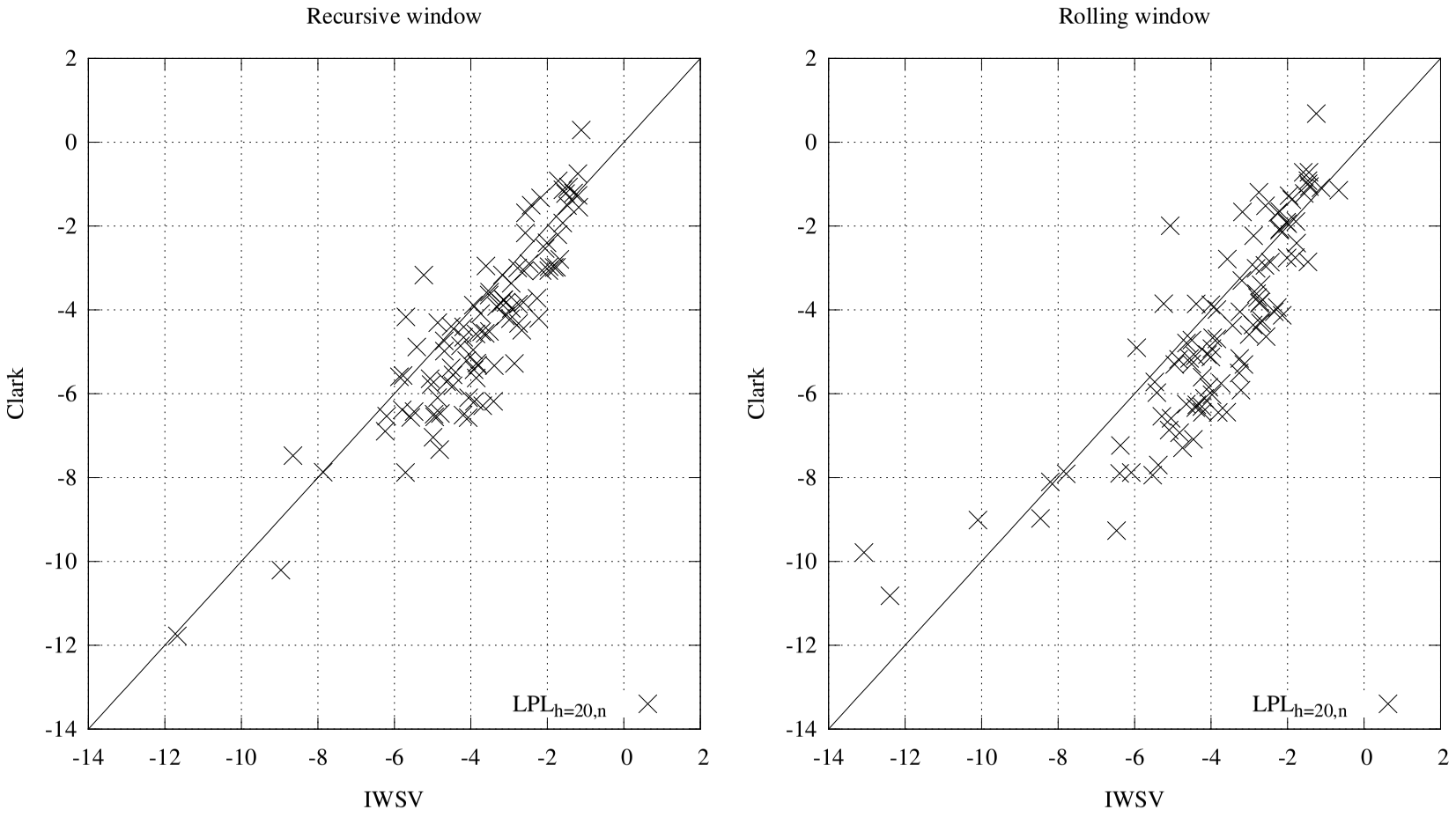


Figure 21.i: GDP growth series $y_{1,t}$ and forecast, $IWSV$ and $Clark$ models for $\mathbf{v_t}$, $n = 100$, Recursive sample window

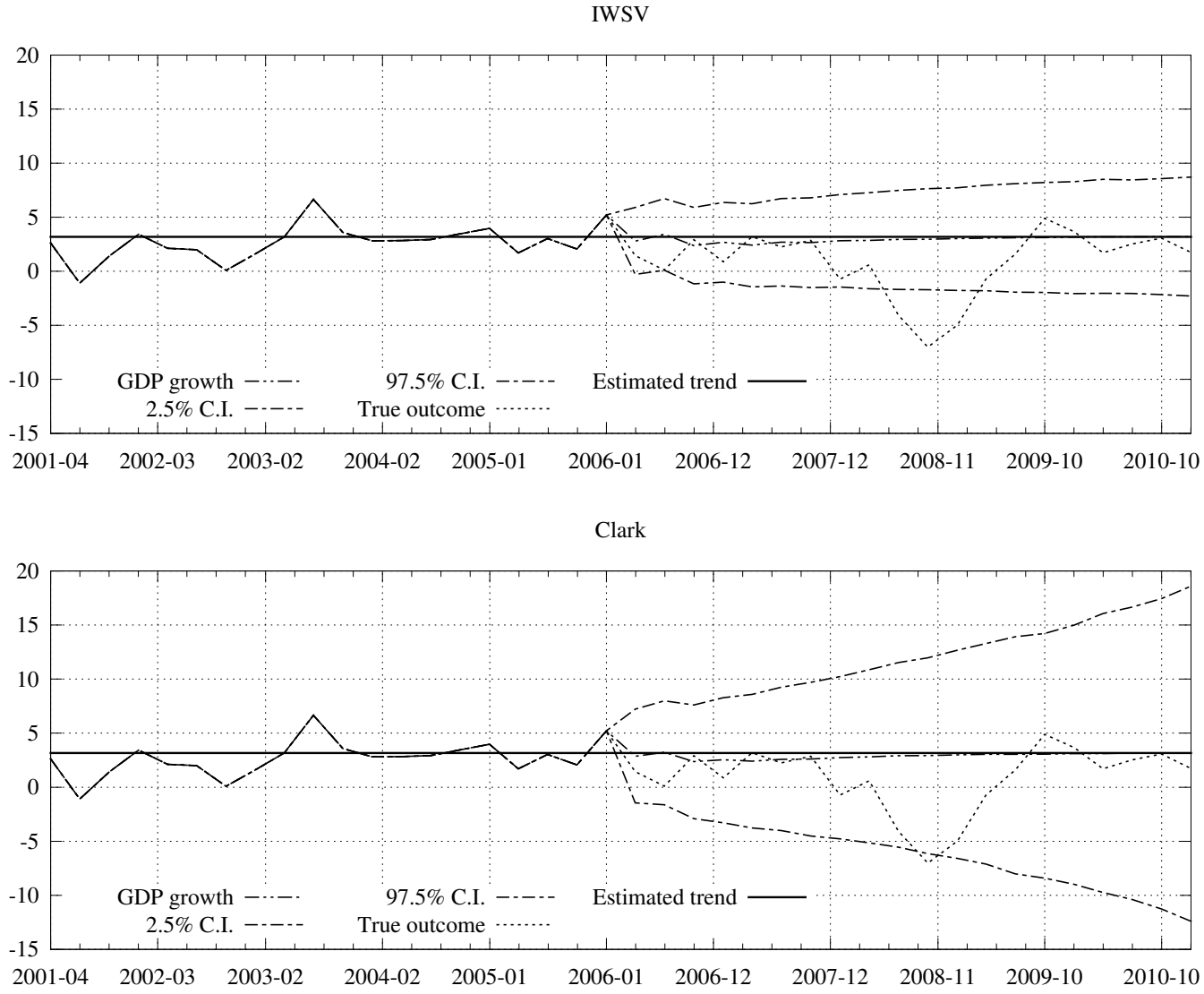


Figure 21.ii: Inflation growth series $y_{2,t}$ and forecast, $IWSV$ and $Clark$ models for $\mathbf{v_t}$, $n = 100$, Recursive sample window

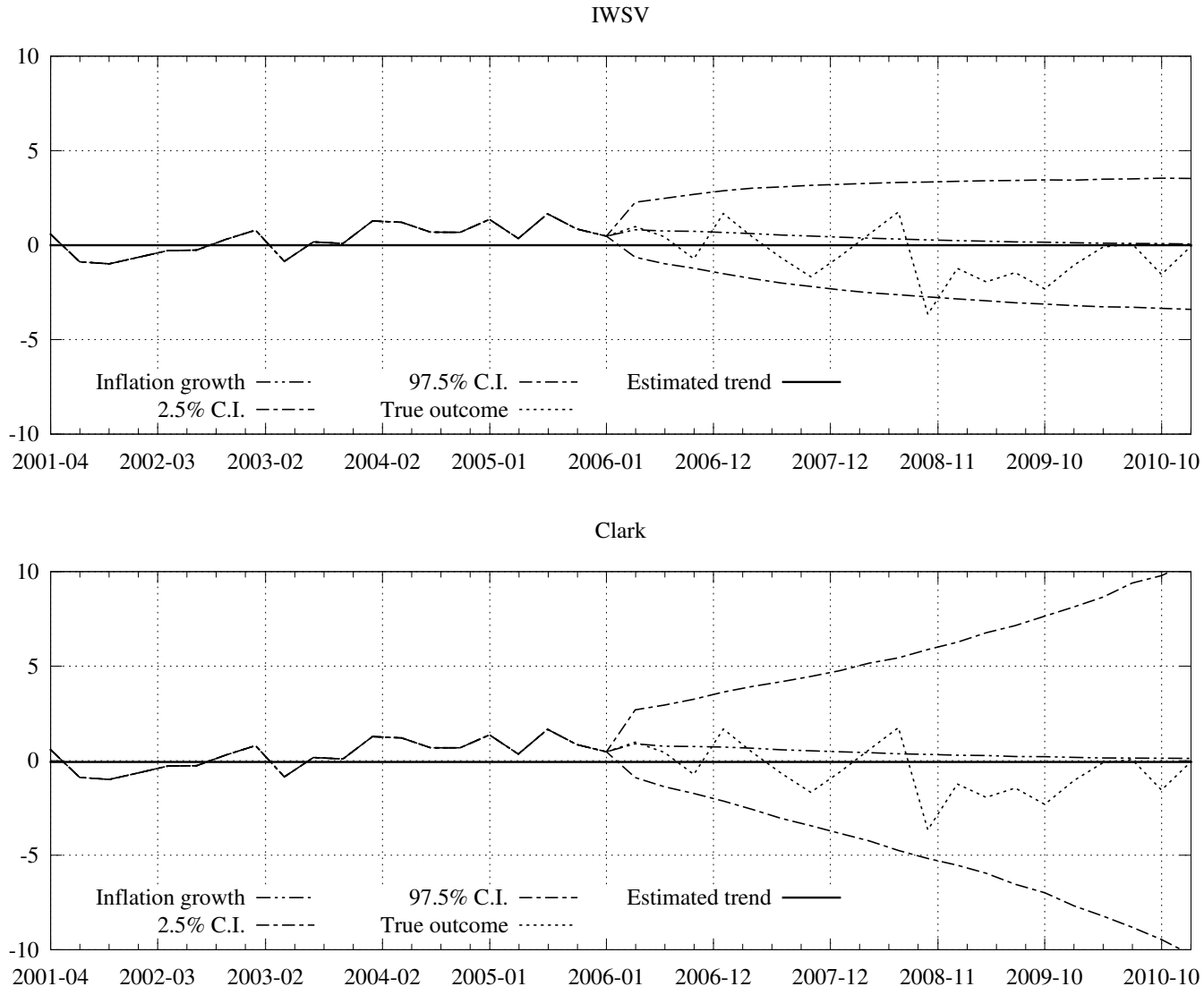


Figure 21.iii: Interest rate series $y_{3,t}$ and forecast, $IWSV$ and $Clark$ models for $\mathbf{v_t}$, $n = 100$, Recursive sample window

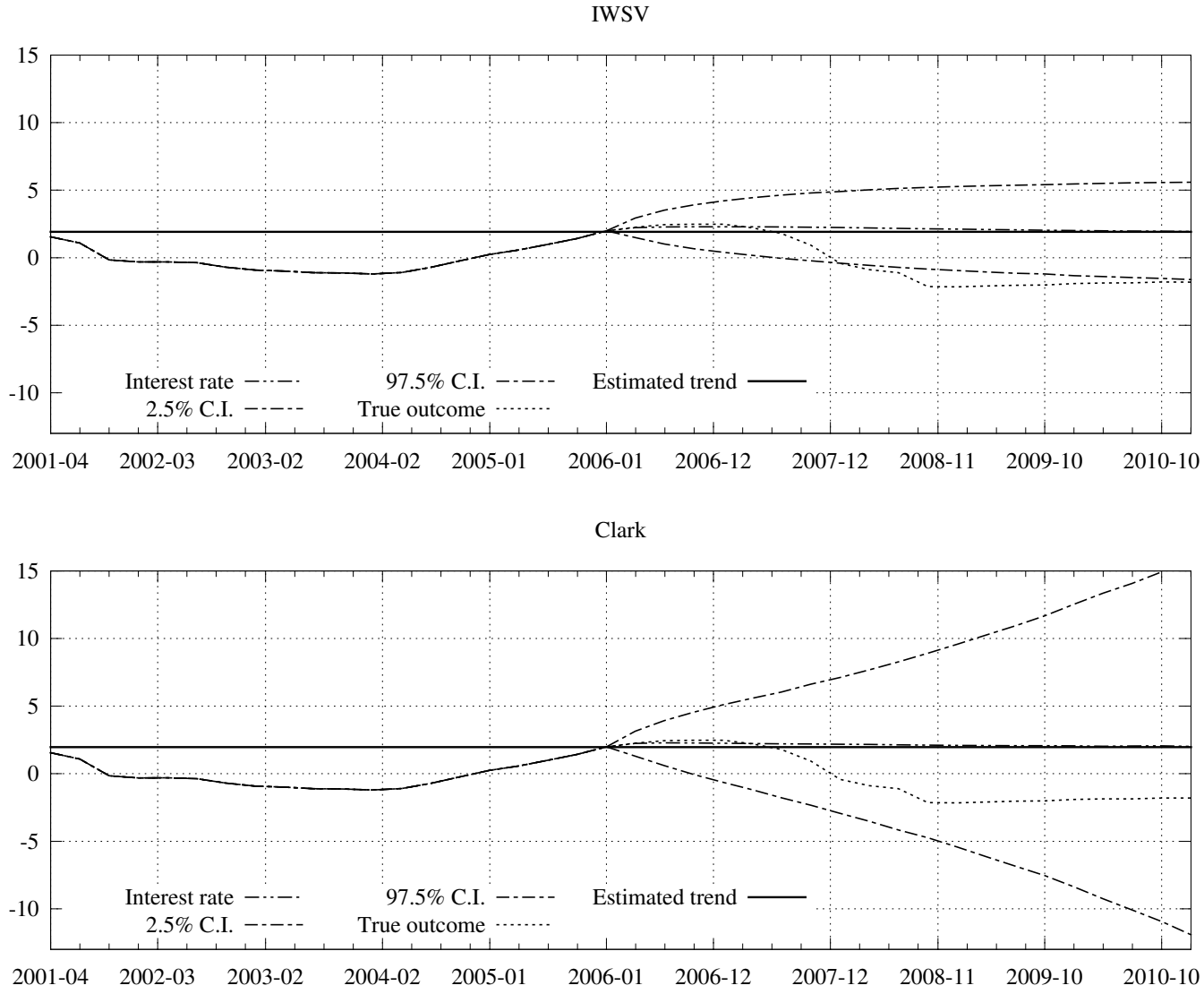
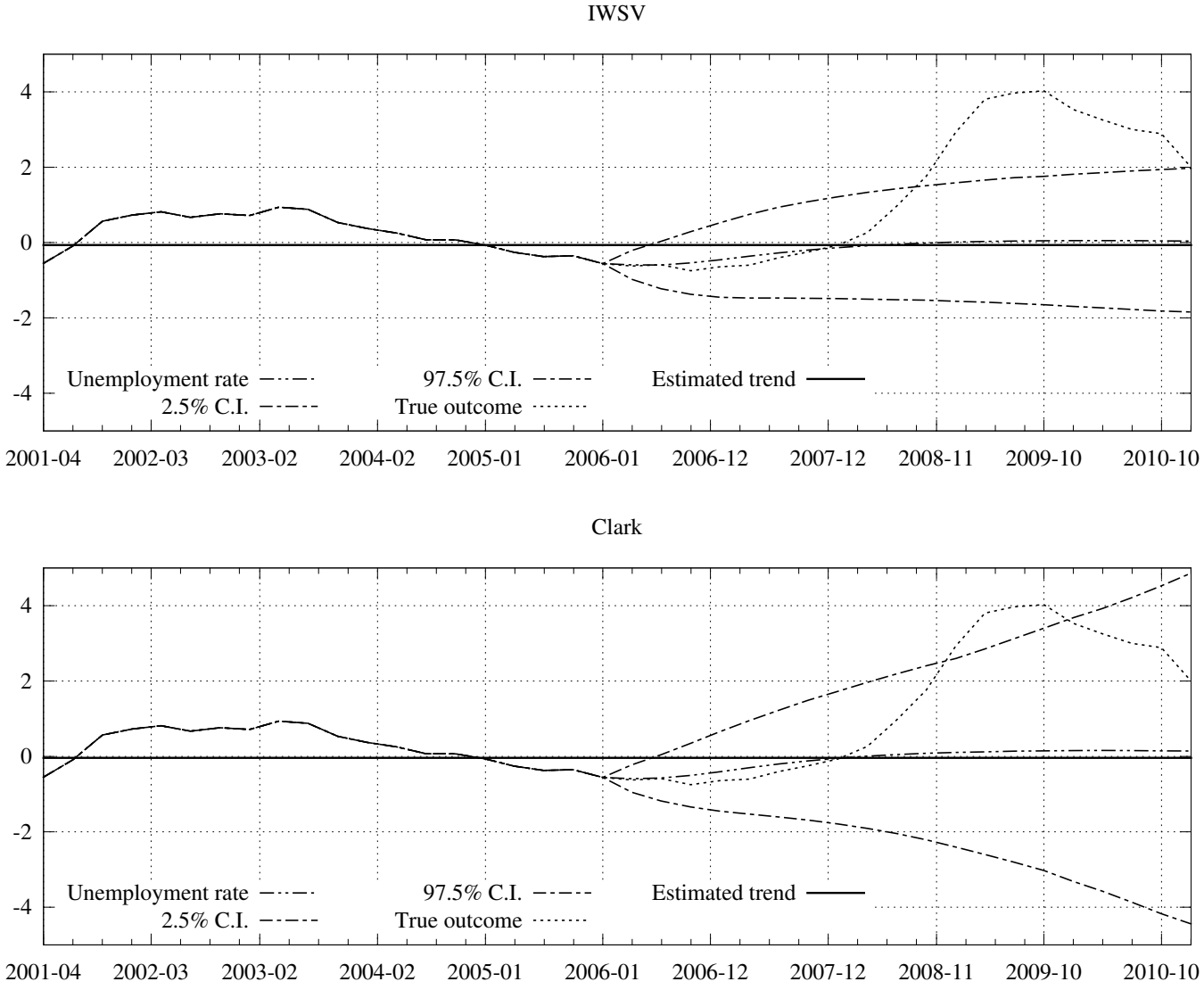


Figure 21.iv: Unemployment rate series $y_{4,t}$ and forecast, $IWSV$ and $Clark$ models for $\mathbf{v_t}$, $n = 100$, Recursive sample window



8 Conclusion

Dramatic changes in macroeconomic time series volatility have posed a challenge to contemporary VAR forecasting models. Traditionally, the conditional volatility of these models had been assumed constant over time or allowed for structural breaks across long time periods. More recent work, however, has improved forecasts by allowing the conditional volatility to be completely time variant by specifying the VAR innovation variance as a distinct discrete time process. For example, Clark (2011) specified the elements of the covariance matrix process of the VAR innovations as linear functions of independent nonstationary processes.

However, it is not clear that the choice of nonstationary driving processes is suitable. Moreover, in order to reduce parameterization, some form of fixed relationship is imposed between the elements of the VAR innovation covariance matrix and the independent processes driving them.

Ultimately, we would like to have an empirical rationale for this choice of specification. Given this, we have proposed and tested both the Clark (2011) benchmark model, and the alternative multivariate volatility process, $IWSV$, which is constructed in such a way to directly model the time varying covariance matrices by means of the Inverse Wishart distribution. These models have been estimated, and forecasts been constructed, on a data set as close to Clark (2011) as possible.

Motivating this study are also a number of theoretical advantages of the proposed $IWSV$ specification. For one, the direct specification of the dynamics of the latent stochastic volatility process, $\mathbf{\Sigma_t}$, precludes the need to specify convoluted relationships between the (co)volatility elements of $\mathbf{\Gamma_t}$, $\gamma_{ij,t}$, and the driving processes $\lambda_{i,t}$, through the $\mathbf{B}$ matrix. Moreover, the autoregressive dynamics between volatility series are more

easily interpreted as volatility spill-over effects, since we no longer need to disentangle these relationships. The model is now also invariant to permutation of the order of the observed series. Finally, it is easy to derive conditions ensuring the existence of the unconditional mean of the processes $(\mathbf{\Sigma_t})$ and $(\mathbf{y_t})$.

In applying both models to the data, we have chosen to evaluate their performance strictly in terms of forecasting ability. In doing so, we have chosen to evaluate both point and interval forecasts. Point forecasts consider the mean squared error (MSE) of out-of-sample forecasts, while interval forecasts considered the Bayesian log predictive likelihood measure, $LPL_{h,n}$, along a number of forecast horizons. Moreover, we have computed these metrics a number of times, across sample windows, which represent subsamples of the entire data set.

Estimating both models provides a number of interesting results. First, posterior means of the parameters are much more stable across the sample windows, given a recursive window that grows larger, than a rolling window of fixed size. Moreover, irrespective of sample window size or scheme, the posterior means of many parameters deviate from our assumptions on the prior means, suggesting that the data are informative, despite the small sample size.

Interestingly, the stationarity of the multivariate volatility process driving the VAR innovations may be questionable, as the $IWSV$ model estimates seem to suggest the possibility of at least one unit root. We also find that the filtered latent (co)volatilities of the $Clark$ model are much too smooth and suggest too little variation across time.

Forecasting performance is also mixed. For example, the MSE out-of-sample forecasts suggest that neither model exhibits a strong advantage. Turning to interval forecasts, we consider the distribution of the log predictive likelihood, $LPL_{h,n}$, measures for both models. While the $IWSV$ exhibits a strong advantage in the sense that the

mean of these measures are substantially improved, this model suffers from the fact that the measures are more skewed in the negative direction, and prone to rare occurrences of dramatically poor forecast fit, with this sensitivity becoming worse as the forecast horizon grows large. Finally, as expected, given the dramatic nonstationarity of the $Clark$ volatility process, its prediction intervals tend to grow exponentially.

Ultimately, we must emphasize that our methodology does not presume that one of the competing models is well specified, instead we insist on the opposite. Rather, given these results we suggest an approach that might make use jointly of both the $Clark$ and $IWSV$ specifications in practice. That is, use, say, the $Clark$ in some environments and the $IWSV$ in others. Moreover, we could likely encompass both models in a better, also misspecified, model. A natural idea is to introduce a model with endogenous switching regimes, where the $Clark$ model is employed in situations where it performs better, while the $IWSV$ can be used otherwise. It is important to note that this encompassing model would not represent a mixture of the $Clark$ and $IWSV$ specifications, with unknown mixing weights, regularly updated by Bayesian techniques.

9 References

BERNANKE, B., AND I. MIHOV (1998a): "The Liquidity Effect and Long-Run Neutrality," *Carnegie-Rochester Conference Series on Public Policy*, 49, 149-194.

————————- (1998b): "Measuring Monetary Policy," *Quarterly Journal of Economics*, 113, 869-902.

CHIB, S., Y. OMORI, AND M. ASAI (2009): "Multivariate Stochastic Volatility," in *Handbook of Financial Time Series.,* ed. by T.G. Anderson, et al., Berlin: Springer-Verlag Publishing.

CLARK, T.E. (2011): "Real-Time Density Forecasts from Bayesian Vector Autoregressions with Stochastic Volatility," *Journal of Business & Economic Statistics,* 29, 3, 327-341.

CLARK, T.E., AND M.W. McCRACKEN (2001): "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics,* 105, 85-110.

————————- (2008): "Forecasting with Small Macroeconomic VARs in the Presence of Instability", in *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, ed. by D.E. Rapach and M.E. Wohar, Bingley, UK: Emerald Publishing.

————————- (2010): "Averaging Forecasts from VARs with Uncertain Instabilities", Journal of Applied Econometrics, 25, 5-29.

CLEMENTS, M.P., AND D.F. HENDRY (1998): *Forecasting Economic Time Series,* Cambridge, U.K.: Cambridge University Press.

COGLEY, T., AND T.J. SARGENT (2001): "Evolving Post-World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual*, 16, 331-373.

————————- (2005) "Drifts and Volatilities: Monetary Policies and Outcomes in the

Post-World War II U.S.," *Review of Economic Dynamics*, 8, 262-302.

CROUSHORE, D. (2006): "Forecasting with Real-Time Macroeconomic Data," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, Amsterdam: North-Holland Publishers.

DIEBOLD, F.X., AND K. YILMAZ (2008): "Measuring Financial Asset Return and Volatility Spillovers with Application to Global Equity Markets," *Working Paper 08-16, Research Department, Federal Reserve Bank of Philadelphia*.

ENGLE, R.F., AND K.F. KRONER (1995): "Multivariate Simultaneous Generalized ARCH," *Economic Theory*, 11, 122-150.

FOX, E.B., AND M. WEST (2011): "Autoregressive Models for Variance Matrices: Stationary Inverse Wishart Processes," *arXiv:*1107.5239v1.

GEWEKE, J., AND G. AMISANO (2010): "Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns," *International Journal of Forecasting*, 26, 216-230.

GOURIEROUX, C., J. JASIAK, AND R. SUFANA (2009): "The Wishart Autoregressive Process of Multivariate Stochastic Volatility," *Journal of Econometrics*, 150, 167-181.

GOLOSNOY, V., B. GRIBISCH, AND R. LIESENFELD (2010): "The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility," *Working Paper, Christian-Albrechts-Universitat zu Kiel*, 07.

GREENBURG, E. (2008): *Introduction to Bayesian Econometrics*, Cambridge, U.K.: Cambridge University Press.

HAMILTON, J.D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica,* 57, 357-384.

JORE, A.S., J. MITCHELL, AND S.P. VAHEY (2010): "Combining Forecast Densities from VARs with Uncertain Instabilities," *Journal of Applied Econometrics*, 25, 621-634.

KIM, C.J., AND C.R. NELSON (1999): "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov Switching Model of the Business Cycle," *Review of Economics and Statistics,* 81, 608-661.

KOZICKI, S., AND P.A. TINSLEY (2001a): "Shifting Endpoints in the Term Structure of Interest Rates," *Journal of Monetary Economics*, 47, 613-652.

————————- (2001b): "Term Structure Views of Monetary Policy Under Alternative Models of Agent Expectations," *Journal of Economic Dynamics and Control*, 25, 149-184.

LITTERMAN, R.B. (1986): "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25-38.

MAGNUS, J.R., AND H. NEUDECKER (1980): "The Elimination Matrix: Some Lemmas and Applications," *SIAM Journal on Algebraic and Discrete Methods*, 1, 422-449.

————————- (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: John Wiley and Sons Publishers.

McCONNELL, M., AND G. PEREZ QUIROS (2000): "Output Fluctuations in the United States: What Has Changed Since the Early 1980s?," *American Economic Review*, 90, 1464-1476.

PHILIPOV, A., AND M.E. GLICKMAN (2006): "Multivariate Stochastic Volatility via Wishart Processes," *Journal of Business and Economic Statistics*, 24, 3, 313-328.

PRESS, S.J. (1982): *Applied Multivariate Analysis*, 2nd ed., New York: Dover Publications.

PRIMICERI, G. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821-852.

RINNERGSCHWENTNER, W., G. TAPPEINER, AND J. WALDE (2011): "Multivariate Stochastic Volatility via Wishart Processes – A Continuation," *Working Papers in Economics and Statistics: University of Innsbruck*, 19.

ROMER, C.D., AND D.H. ROMER (2000): "Federal Reserve Information and the Behaviour of Interest Rates," *American Economic Review*, 90, 429-457.

SARTORE, D., AND M. BILLIO (2005): "Stochastic Volatility Models: A Survey with Applications to Option Pricing and Value at Risk," in *Applied Quantitative Methods for Trading and Investment,* ed. by C.L. Dunis, J. Laws, and P. Naim, John Wiley and Sons Publishers.

SIMS, C.A. (2001): "Comment on Sargent and Cogley's 'Evolving Post-World War II U.S. Inflation Dynamics'," *NBER Macroeconomics Annual*, 16, 373-379.

————— (2002): "The Role of Models and Probabilities in the Monetary Policy Process," *Brookings Papers on Economic Activity,* 2, 1-40.

STOCK, J.H. (2001): "Discussion of Cogley and Sargent 'Evolving Post-World War II U.S. Inflation Dynamics'," *NBER Macroeconomics Annual*, 16, 379-387.

VILLANI, M. (2009): "Steady-State Priors for Vector Autoregressions," *Journal of Applied Econometrics*, 24, 630-650.

10 Appendix: Real data, $LDL^{'}$ factorization of $\mathbf{\Sigma_t}$

Further evidence for the $IWSV$ model can be drawn by considering the form of the constraint which the $Clark$ model places on the time varying structure of the covariance matrix process of the VAR innovations, $\mathbf{v_t}$. Recall from Section 3.1 that the $Clark$ model imposes the following parameterization on the VAR innovations given as

$$\mathbf{v_t}=\mathbf{B}^{-1}\mathbf{\Lambda}_t^{0.5}\varepsilon_{\mathbf{t}}, \quad \text{where} \quad \varepsilon_{\mathbf{t}} \sim MVN_p\left(\mathbf{0}, \mathbf{I_p}\right). \tag{47}$$

Of course, the constraint above implies that

$$\mathbf{\Gamma_t}=\mathbf{B}^{-1}\mathbf{\Lambda_t}\left(\mathbf{B}^{-1}\right)^{'} = var\left(\mathbf{v_t}\right). \tag{48}$$

The interesting point to note is that this parameterization is equivalent to imposing an $LDL^{'}$ factorization on the covariance matrix of the VAR innovations, where L is a lower triangular matrix with ones on the diagonal and D is a diagonal matrix. Note that this $LDL^{'}$ factorization always exists for positive definite, real, symmetric matrices and is unique.

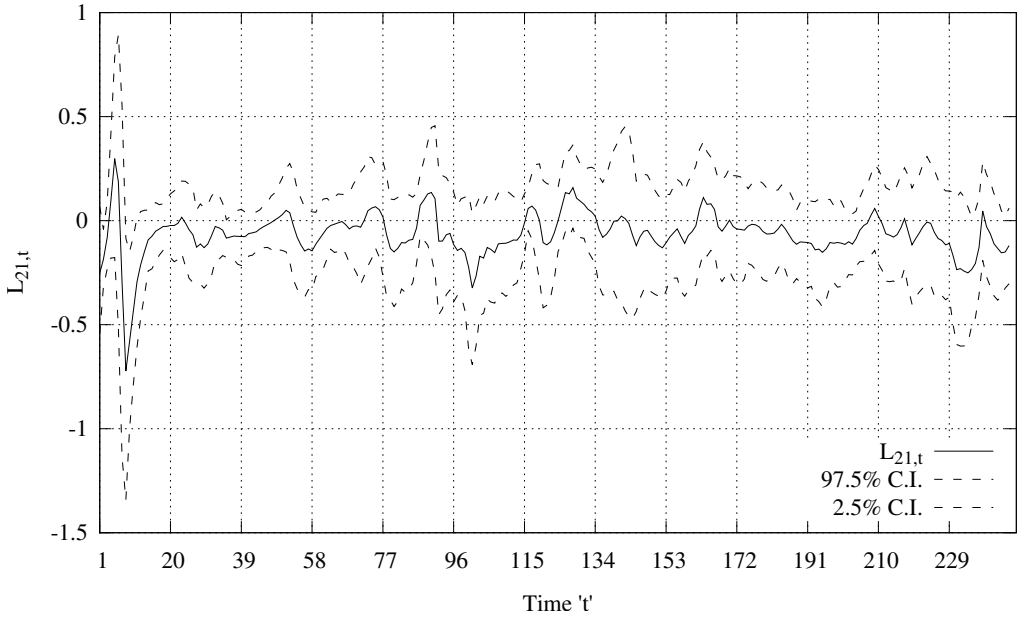
This result implies a method for testing whether Clark's parametric assumption on the volatility process is correct. Suppose that we estimate the VAR model under the $IWSV$ specification, given the entire data set. At each point in time that we draw a covariance matrix $\mathbf{\Sigma_t}^{(m)}$, from the Gibbs sampler, we factorize this covariance matrix as $\mathbf{\Sigma}_t = L_t D_t L_t^{'}$. Iterating in this way provides us with a finite sample distribution of the $L_t^{(m)}$ matrices implied by the $IWSV$ specification for each time period $t = 1, \ldots, T$. Since these L_t matrices are unique, their time varying distributions must suggest something about whether or not it is appropriate to assume that the elements of the $\mathbf{B}^{-1}$ matrix in the $Clark$ specification should be specified as constant across time.

Of course, while there are other ways to check the validity of this assumption (e.g. estimate the $Clark$ with time varying $\mathbf{B}$ matrices and compare results according to some metric), the aforementioned test proves the most immediately applicable.

Figures 22.i and **22.ii** illustrate the results of this test which suggest that there exists significant time variation in the elements of the L_t matrix factors across time, especially with regards to the elements corresponding to the pairing of GDP growth with both the

inflation rate and the interest rate (i.e. $L_{21,t}$ and $L_{31,t}$).

Figure 22.i: $L_{21,t}$, time varying density

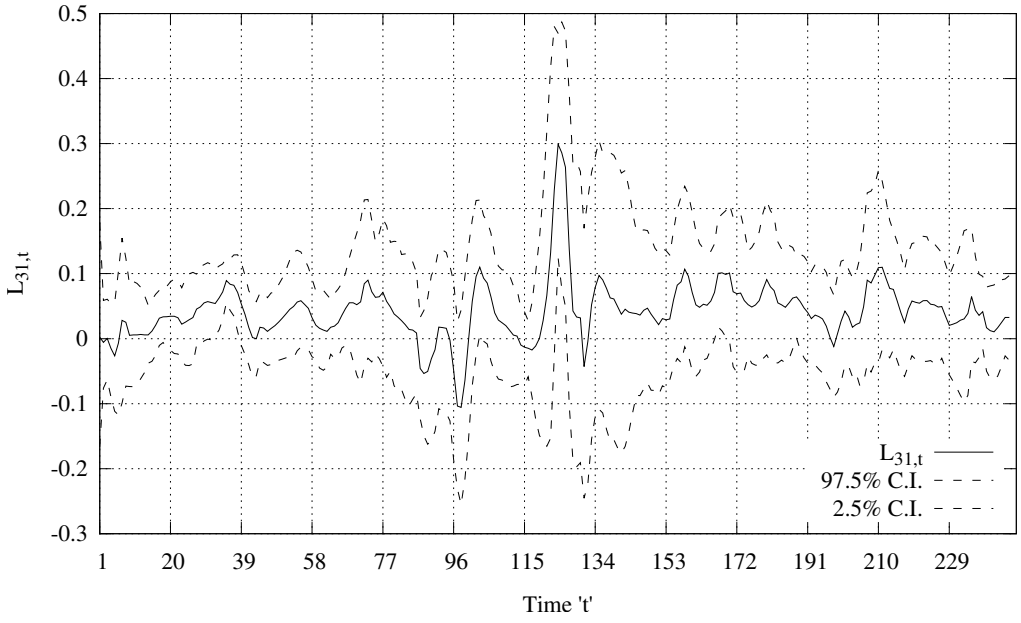


11 Appendix: Derivation of the posterior distributions needed for Gibbs sampling

The following appendix describes the steps required for generating the conditional posterior distribution of the parameters used in the Gibbs sampler. First, we consider the steps required for estimation of the $VAR(J)$ model with $Clark$ volatility, then we consider the steps required for estimating the $IWSV$ parameters. Please refer to Section 5 for a summary of the steps outlined below.

The benchmark volatility specification, $Clark$, is estimated by Gibbs sampler, given estimation steps based on those of Villani (2009) and Cogley and Sargent (2005). Clark (2011) provides expressions for the conditional posterior distributions of the VAR parameters of his Gibbs sampler, obtained from Mattias Villani, who himself derived them based on the constant variance sampler employed in Villani (2009). We have re-derived them here for completeness.

Figure 22.ii: $L_{31,t}$, time varying density



11.1 Definition of the parameters and priors

What follows provides descriptions of the model parameters and their chosen prior distributions.

i) Parameters related to the $VAR(J)$ process:

- $\mathbf{\Pi}_j$, for $j = 1, \ldots, J$, the autoregressive coefficient matrices on the $VAR(J)$ specification. Each $\mathbf{\Pi}_i$ is of dimension $p \times p$. Combining all the J coefficient matrices, $\mathbf{\Pi_j}$, into one larger matrix, $\mathbf{\Pi}$, the conditionally conjugate prior is multivariate Normal, $\mathbf{\Pi} \sim N(\mu_{\mathbf{\Pi}}, \mathbf{\Xi_\Pi})$.

- $\mathbf{\Psi}$, the matrix which when multiplied by the deterministic trend vector $\mathbf{d_t}$, forms the unconditional mean vector of the VAR specification. $\mathbf{\Psi}$ is of dimension $p \times q$. The conditionally conjugate prior is multivariate Normal, $\mathbf{\Psi} \sim N(\mu_{\mathbf{\Psi}}, \mathbf{\Xi_\Psi})$.

ii) (Augmented) parameters specific to the Clark (2011) volatility specification:

- $\mathbf{B}$, the lower triangular matrix with ones along its main diagonal, which when inverted, and pre and post multiplied by $\mathbf{\Lambda_t}$, forms the covariance matrix of the

VAR shocks, v_t, given as $\mathbf{\Gamma_t}$ below. $\mathbf{B}$ is of dimension $p \times p$. The elements of the $\mathbf{B}$ matrix are assumed independent Normal, with details provided in the relevant section below.

- $\mathbf{\Lambda}_t$, the diagonal matrix which contains the nonstationary, independent, driving processes $\lambda_{i,t}$, for $i = 1, \ldots, p$, along its main diagonal. $\mathbf{\Lambda}_t$ is of dimension $p \times p$. The prior for this augmented parameter is Log-normal with details provided in the relevant section below.

- $\mathbf{\Phi}$, the diagonal matrix which contains the variances, φ_i, of the shocks of the driving processes, ξ_i, for $i = 1, \ldots, p$. $\mathbf{\Phi}$ is of dimension $p \times p$. We assume conditionally conjugate, independent Inverse-Gamma priors on each $\varphi_i \sim IG(\frac{\gamma}{2}, \frac{\delta}{2})$.

iii) (Augmented) parameters specific to the proposed IWSV volatility specification:

- $\mathbf{A}_k$, for $k = 1, \ldots, K$, the autoregressive coefficient matrices on the $IWSV$ volatility specification. Each $\mathbf{A}_k$ is of dimension $p \times p$, and is not necessarily symmetric. We assume multivariate Normal priors for each of the $\mathbf{A}_k$, for $k = 1, \ldots, K$, matrices.

- $\mathbf{C}$ the lower diagonal constant matrix in the $IWSV$ volatility specification. $\mathbf{C}$ is of dimension $p \times p$. We assume a multivariate Normal prior for on the $\mathbf{C}$ matrix.

- ν the degree of freedom parameter describing the shape of the Inverse Wishart distribution driving the volatility shocks. The degree of freedom parameter is a scalar. The prior chosen is a Gamma distribution.

- $\mathbf{\Sigma}_t$, $\forall t$, the covariance matrices of the VAR process shocks under the $IWSV$ model. The $\mathbf{\Sigma}_t$ matrices are all of dimension $p \times p$. The prior is Inverse Wishart with details provided in the relevant section below.

11.2 Computation of the posterior distribution for the Clark (2011) volatility model

Let us now describe the different conditional posterior distributions involved in the sequence for Gibbs sampling.

1. Draw from the posterior density of the slope coefficients $\mathbf{\Pi}^{'} = [\mathbf{\Pi_1}, \mathbf{\Pi_2}, \ldots, \mathbf{\Pi_J}]$ of the VAR, conditional on $\mathbf{\Psi}$, $\underline{\mathbf{\Lambda_T}}$, $\mathbf{B}$, $\mathbf{\Phi}$=$diag(\varphi_1, \varphi_2, \ldots, \varphi_p)$ and the data, given multivariate Normal prior, $\mathbf{\Pi} \sim N(\mu_{\mathbf{\Pi}}, \mathbf{\Omega}_{T\mathbf{\Pi}})$.

   For this step we rewrite the VAR as:

$$\mathbf{Y_t} = \mathbf{\Pi}^{'}\mathbf{X_t} + \mathbf{v_t}, \tag{49a}$$

$$\text{where} \quad \mathbf{Y_t} = \mathbf{y_t} - \mathbf{\Psi d_t}, \tag{49b}$$

$$\mathbf{v_t} = \mathbf{B}^{-1}\mathbf{\Lambda}_t^{0.5}\varepsilon_{\mathbf{t}}, \tag{49c}$$

$$\text{and} \quad \mathbf{X_t} = \left[(\mathbf{y_{t-1}} - \mathbf{\Psi d_{t-1}})^{'}, (\mathbf{y_{t-2}} - \mathbf{\Psi d_{t-2}})^{'}, \ldots, (\mathbf{y_{t-J}} - \mathbf{\Psi d_{t-J}})^{'}\right]^{'}. \tag{49d}$$

   This is a linear model with respect to the parameter elements of the matrix $\mathbf{\Pi}$. To clearly illustrate this linear model we can use the alternative expression [Magnus and Neudecker (1998)]:

$$\begin{aligned}\mathbf{Y_t} &= vec\left(\mathbf{\Pi}^{'}\mathbf{X_t}\right) + \mathbf{v_t} \\ &= \left(\mathbf{I_p} \otimes \mathbf{X_t}^{'}\right) \cdot vec\left(\mathbf{\Pi}\right) + \mathbf{v_t},\end{aligned} \tag{50a}$$

   where $\otimes$ denotes the tensor product. Eliminating the heteroskedasticity by pre-multiplication we have

$$\mathbf{Y_t^*} = \mathbf{\Gamma_t^{-0.5}}\mathbf{Y_t} = \mathbf{\Gamma_t^{-0.5}}\left(\mathbf{I_p} \otimes \mathbf{X_t}^{'}\right) \cdot vec\left(\mathbf{\Pi}\right) + \varepsilon_{\mathbf{t}}, \tag{51}$$

   where $\varepsilon_{\mathbf{t}} \sim N\left(\mathbf{0}, \mathbf{I_p}\right)$ and $\mathbf{\Gamma_t^{-0.5}} = \mathbf{\Lambda_t^{-0.5}}\mathbf{B}$. Or equivalently with clear notation:

$$\mathbf{Y_t^*} = \mathbf{X_t^*} vec(\mathbf{\Pi}) + \varepsilon_{\mathbf{t}}, \tag{52}$$

   where $\mathbf{X_t^*} = \mathbf{\Gamma_t^{-0.5}}\left(\mathbf{I_p} \otimes \mathbf{X_t}^{'}\right)$.

   Thus we have the following Lemma [See Tsay (2005), Section 12.3.2, or Box and Tiao (1973)].

   Lemma 11.1.

Let us consider the Gaussian regression model

$$\mathbf{Y}_{\mathbf{t}}^{*} = \mathbf{X}_{\mathbf{t}}^{*} vec(\mathbf{\Pi}) + \varepsilon_{\mathbf{t}}, \tag{53}$$

with prior,

$$vec(\mathbf{\Pi}) \sim N(\mu_{\mathbf{\Pi}}, \mathbf{\Xi}_{\mathbf{\Pi}}). \tag{54}$$

Then the posterior distribution is such that $vec(\mathbf{\Pi}) \sim N(\mu_{\mathbf{\Pi}}^{*}, \mathbf{\Xi}_{\mathbf{\Pi}}^{*})$*, where*

$$\mathbf{\Xi}_{\mathbf{\Pi}}^{*} = \left[\mathbf{\Xi}_{\mathbf{\Pi}}^{-1} + \mathbf{X}^{*'}\mathbf{X}^{*}\right]^{-\mathbf{1}} \tag{55a}$$

$$\textit{and} \quad \mu_{\mathbf{\Pi}}^{*} = \mathbf{\Xi}_{\mathbf{\Pi}}^{*}[\mathbf{\Xi}_{\mathbf{\Pi}}^{-1}\mu_{\mathbf{\Pi}} + \mathbf{X}^{*'}\mathbf{Y}^{*}], \tag{55b}$$

$$\tag{55c}$$

given $\mathbf{X}^{*}\mathbf{X}^{*'} = \sum_{t=1}^{T} \mathbf{X}_{\mathbf{t}}^{*'}\mathbf{X}_{\mathbf{t}}^{*}$ *and* $\mathbf{X}^{*}\mathbf{Y}^{*'} = \sum_{t=1}^{T} \mathbf{X}_{\mathbf{t}}^{*'}\mathbf{Y}_{\mathbf{t}}^{*}$.

The sufficient summaries of the past, appearing in this posterior distribution, are given as

$$\mathbf{X}^{*'}\mathbf{X}^{*} = \sum_{t=1}^{T} \left[\mathbf{\Gamma}_{\mathbf{t}}^{-0.5}\left(\mathbf{I}_{p} \otimes \mathbf{X}_{\mathbf{t}}^{'}\right)\right]^{'} \mathbf{\Gamma}_{\mathbf{t}}^{-0.5}\left(\mathbf{I}_{\mathbf{p}} \otimes \mathbf{X}_{\mathbf{t}}^{'}\right) \tag{56a}$$

$$= \sum_{t=1}^{T} \left(\mathbf{I}_{p} \otimes \mathbf{X}_{\mathbf{t}}^{'}\right)^{'} \mathbf{\Gamma}_{\mathbf{t}}^{-1}\left(\mathbf{I}_{\mathbf{p}} \otimes \mathbf{X}_{\mathbf{t}}^{'}\right) \tag{56b}$$

$$\text{and} \quad \mathbf{X}^{*'}\mathbf{Y}^{*} = \sum_{\mathbf{t}=1}^{T} \left[\mathbf{\Gamma}_{\mathbf{t}}^{-0.5}\left(\mathbf{I}_{\mathbf{p}} \otimes \mathbf{X}_{\mathbf{t}}^{'}\right)\right]^{'} (\mathbf{\Gamma}_{\mathbf{t}}^{-0.5}\mathbf{Y}_{\mathbf{t}}) \tag{56c}$$

$$= \sum_{t=1}^{T} \left(\mathbf{I}_{p} \otimes \mathbf{X}_{\mathbf{t}}^{'}\right)^{'} \mathbf{\Gamma}_{\mathbf{t}}^{-1}\mathbf{Y}_{\mathbf{t}}. \tag{56d}$$

2. Draw from the posterior density of the coefficients $\mathbf{\Psi}$ defining the trend, conditional on $\mathbf{\Pi}$, $\underline{\mathbf{\Lambda}_{\mathbf{T}}}$, $\mathbf{B}$, and $\mathbf{\Phi}$=$diag(\varphi_1, \varphi_2, \ldots, \varphi_p)$ and the data, given a multivariate Normal prior, $\mathbf{\Psi} \sim N(\mu_{\mathbf{\Psi}}, \mathbf{\Xi}_{\mathbf{\Psi}})$.

The equation defining $\mathbf{Y_t}$ can be rewritten as:

$$\mathbf{Y_t} = \mathbf{\Pi}\left(L\right)\mathbf{y_t} = \mathbf{\Pi}\left(L\right)\mathbf{\Psi d_t} + \mathbf{v_t}. \tag{57}$$

Let us check that this is still a linear model. We have

$$\mathbf{Y_t} = \left(\mathbf{I_p} - \sum_{j=1}^{J}\mathbf{\Pi}_j L^j\right)\mathbf{\Psi d_t} + v_t \tag{58a}$$

$$= \mathbf{I_p\Psi d_t} - \mathbf{\Pi}_1\mathbf{\Psi d_{t-1}} - \cdots - \mathbf{\Pi}_J\mathbf{\Psi d_{t-J}} + v_t \tag{58b}$$

$$= \left(\mathbf{d_t}' \otimes \mathbf{I_p}\right)\cdot vec(\mathbf{\Psi}) - \left(\mathbf{d_{t-1}}' \otimes \mathbf{\Pi}_1\right)\cdot vec(\mathbf{\Psi}) - \cdots - \left(\mathbf{d_{t-J}}' \otimes \mathbf{\Pi}_J\right)\cdot vec(\mathbf{\Psi}) + v_t \tag{58c}$$

$$= \left(\left(\mathbf{d_t}' \otimes \mathbf{I_p}\right) - \left(\mathbf{d_{t-1}}' \otimes \mathbf{\Pi}_1\right) - \cdots - \left(\mathbf{d_{t-J}}' \otimes \mathbf{\Pi}_J\right)\right)\cdot vec(\mathbf{\Psi}) + v_t \tag{58d}$$

$$= \mathbf{X_t}\cdot vec(\mathbf{\Psi}) + v_t, \tag{58e}$$

where $\mathbf{X_t} = \left(\left(\mathbf{d_t}' \otimes \mathbf{I_p}\right) - \left(\mathbf{d_{t-1}}' \otimes \mathbf{\Pi}_1\right) - \cdots - \left(\mathbf{d_{t-J}}' \otimes \mathbf{\Pi}_J\right)\right)$.

Thus we can standardize by pre-multiplication and get

$$\mathbf{Y_t^*} = \mathbf{\Gamma_t^{-0.5}Y_t} = \mathbf{\Gamma_t^{-0.5}}\left(\mathbf{\Pi}\left(L\right)\mathbf{y_t}\right) = \mathbf{\Gamma_t^{-0.5}X_t}\cdot vec\left(\mathbf{\Psi}\right) + \varepsilon_\mathbf{t} \tag{59a}$$

$$\equiv \mathbf{X_t^*}vec(\mathbf{\Psi}) + \varepsilon_\mathbf{t}, \quad \text{say.} \tag{59b}$$

We can reapply **Lemma 11.1**, with this new set of explanatory variables, to get

the posterior mean and variance for $vec(\mathbf{\Psi}) \sim N(\mu^*_{\mathbf{\Psi}}, \mathbf{\Xi}^*_{\mathbf{\Psi}})$ as:

$$\mathbf{\Xi}^*_{\mathbf{\Psi}} = \left[\mathbf{\Xi}^{-1}_{\mathbf{\Psi}} + \mathbf{X}^{*'}\mathbf{X}^*\right]^{-\mathbf{1}} \tag{60a}$$

$$\text{and} \quad \mu^*_{\mathbf{\Psi}} = \mathbf{\Xi}^*_{\mathbf{\Psi}}[\mathbf{\Xi}^{-1}_{\mathbf{\Psi}}\mu_{\mathbf{\Psi}} + \mathbf{X}^{*'}\mathbf{Y}^*]. \tag{60b}$$

The sufficient summaries of the past are now given as

$$\mathbf{X}^{*'}\mathbf{X}^* = \sum_{t=1}^{T} \left(\mathbf{\Gamma}_{\mathbf{t}}^{-0.5}\mathbf{X_t}\right)' \mathbf{\Gamma}_{\mathbf{t}}^{-0.5}\mathbf{X_t} = \sum_{t=1}^{T} \mathbf{X_t}' \mathbf{\Gamma}_{\mathbf{t}}^{-1}\mathbf{X_t} \tag{61a}$$

$$\text{and} \quad \mathbf{X}^{*'}\mathbf{Y}^* = \sum_{t=\mathbf{1}}^{T} \left(\mathbf{\Gamma}_{\mathbf{t}}^{-0.5}\mathbf{X_t}\right)' \mathbf{\Gamma}_{\mathbf{t}}^{-0.5}\mathbf{Y_t} = \sum_{t=1}^{T} \mathbf{X_t}' \mathbf{\Gamma}_{\mathbf{t}}^{-1}\mathbf{Y_t}. \tag{61b}$$

3. Draw from the posterior density of the elements of $\mathbf{B}$ (lower triangular with ones in the diagonal) conditional on $\mathbf{\Pi}$, $\mathbf{\Psi}$, $\underline{\mathbf{\Lambda_T}}$, $\mathbf{\Phi}$=$diag(\varphi_1, \varphi_2, \ldots, \varphi_p)$ and the data, given Normal, independent, priors on each of the elements of the $\mathbf{B}$ matrix.

   The system defining $\mathbf{Y_t}$ can now be rewritten as

$$\mathbf{B\Pi}\left(L\right)\left(\mathbf{y_t} - \mathbf{\Psi d_t}\right) = \mathbf{BY_t} = \mathbf{\Lambda}_{\mathbf{t}}^{0.5}\varepsilon_{\mathbf{t}} \tag{62}$$

Since $\mathbf{B}$ is lower triangular, this system of equations reduces to the following

$$Y_{1,t} = \lambda_{1,t}^{0.5}\varepsilon_{1,t} \tag{63a}$$

$$Y_{2,t} = -b_{21}Y_{1,t} + \lambda_{2,t}^{0.5}\varepsilon_{2,t} \tag{63b}$$

$$Y_{3,t} = -b_{31}Y_{1,t} - b_{32}Y_{2,t} + \lambda_{3,t}^{0.5}\varepsilon_{3,t} \tag{63c}$$

$$Y_{4,t} = -b_{41}Y_{1,t} - b_{42}Y_{2,t} - b_{43}Y_{3,t} + \lambda_{4,t}^{0.5}\varepsilon_{4,t} \tag{63d}$$

$$\vdots$$

$$Y_{p,t} = -b_{p1}Y_{1,t} - b_{p2}Y_{2,t} - b_{p3}Y_{3,t} - \ldots - b_{p,(p-1)}Y_{(p-1),t} + \lambda_{p,t}^{0.5}\varepsilon_{p,t} \tag{63e}$$

where $Y_{i,t}$ is the ith element of the $p \times 1$ column vector $\mathbf{\Pi}\left(L\right)\left(\mathbf{y_t} - \mathbf{\Psi d_t}\right) = \mathbf{Y_t}$. We can treat each of the $i = 2, \ldots, p$ equations above as linear regressions. Again, pre-multiplication of each of the i equations by $\lambda_{i,t}^{-0.5}$, $\forall t$, removes the heteroskedasticity. Furthermore, given the assumption of independent Normal prior densities, the conditional posterior for each row vector of $\mathbf{B}$ is also Normal: $N\left(\beta_i{}^*, \mathbf{G}_i{}^*\right)$, $\forall i = 2, \ldots, p$, where

$$\mathbf{G}_i^* = \left[\mathbf{G}_i{}^{-1} + \mathbf{X}_i^{*'}\mathbf{X}_i^*\right]^{-1}, \tag{64a}$$

$$\text{and} \quad \beta_i^* = \mathbf{G}_i^*[\mathbf{G}_i^{-1}\beta_i + \mathbf{X}_i^{*'}\mathbf{Y}_i^*], \tag{64b}$$

and

$$\mathbf{Y}_i^* = \left[\lambda_{i,1}^{-0.5} Y_{i,1}, \ldots, \lambda_{i,T}^{-0.5} Y_{i,T}\right]' \tag{65a}$$

$$\text{and} \quad \mathbf{X}_i^* = \begin{bmatrix} -\lambda_{i,1}^{-0.5} Y_{1,1} & -\lambda_{i,1}^{-0.5} Y_{2,1} & \ldots & -\lambda_{i,1}^{-0.5} Y_{i-1,1} \\ \ldots & \ldots & \ldots & \ldots \\ -\lambda_{i,T}^{-0.5} Y_{1,T} & -\lambda_{i,T}^{-0.5} Y_{2,T} & \ldots & -\lambda_{i,T}^{-0.5} Y_{i-1,T} \end{bmatrix}. \tag{65b}$$

4. Draw from the posterior density of the elements of the time varying covariance matrix $\mathbf{\Lambda_t}$ for each time $t = 1, \ldots, T$ in sequence, each conditional on $\mathbf{\Pi}$, $\mathbf{\Psi}$, $\mathbf{B}$, and $\mathbf{\Phi}$=$diag(\varphi_1, \varphi_2, \ldots, \varphi_p)$ and the data.

   Since the stochastic volatilities are independent of each other for all $i = 1, \ldots, p$, we can estimate the corresponding equation separately. In order to do so, we need an expression for the posterior density of each augmented parameter $\lambda_{i,t}$ conditional on everything else, including the entire macroeconomic series values for all $t = 1, \ldots, T$.

   Since each volatility is Markov of order one, we can write for each $i = 1, \ldots, p$

$$\begin{aligned} &g\left(\lambda_{i,t} \mid \lambda_{i,\setminus t}, \varphi_i, \mathbf{Y}_i^*\right) \propto f\left(\mathbf{Y}_i^* \mid \lambda_i\right) g\left(\lambda_{i,t} \mid \lambda_{i,\setminus t}, \varphi_i\right) \propto f\left(y_{i,t}^* \mid \lambda_{i,t}\right) g\left(\lambda_{i,t} \mid \lambda_{i,\setminus t}, \varphi_i\right) \\ &= f\left(y_{i,t}^* \mid \lambda_{i,t}\right) g\left(\lambda_{i,t} \mid \lambda_{i,t-1}, \varphi_i\right) g\left(\lambda_{i,t+1} \mid \lambda_{i,t}, \varphi_i\right) = f\left(y_{i,t}^* \mid \lambda_{i,t}\right) g\left(\lambda_{i,t} \mid \lambda_{i,t-1}, \lambda_{i,t+1}, \varphi_i\right) \end{aligned} \tag{66}$$

   where $\lambda_{i,\setminus t}$ denotes all elements of the λ_i vector except for the tth element, $\mathbf{Y}_i^* = \left\{y_{i,1}^*, \ldots, y_{i,T}^*\right\}$, and $y_{i,t}^*$ is the ith element of $\mathbf{B\Pi}\left(L\right)\left(\mathbf{y_t} - \mathbf{\Psi d_t}\right)$. Furthermore, since

$$\lambda_{i,t} | \lambda_{i,t-1} \sim LN\left(e^{\ln(\lambda_{i,t-1}) + \frac{\varphi_i}{2}}, (e^{\varphi_i} - 1) e^{2\ln(\lambda_{i,t-1}) + \varphi_i}\right), \tag{67}$$

we have

$$f\left(y_{i,t}^{*} \mid \lambda_{i,t}\right) g\left(\lambda_{i,t} \mid \lambda_{i,t-1}, \lambda_{i,t+1}, \varphi_{i}\right)$$
$$\propto \lambda_{i,t}^{-0.5} \exp\left(-\frac{\left(y_{i,t}^{*}\right)^{2}}{2\lambda_{i,t}}\right) \lambda_{i,t}^{-1} \exp\left(-\frac{\left(ln\left(\lambda_{i,t}\right) - \mu_{i,t}\right)^{2}}{2\sigma^{2}}\right), \quad (68)$$

where we can solve for missing values according to Section 12.6.1 of Tsay (2005) and find that

$$\mu_{i,t} = \frac{1}{2}\left(\ln\left(\lambda_{i,t+1}\right) + \ln\left(\lambda_{i,t-1}\right)\right), \quad (69a)$$

$$\text{and} \quad \sigma^{2} = \frac{1}{2}\varphi_{i}. \quad (69b)$$

Therefore, in implementing a Metropolis-within-Gibbs step we can draw a proposal from the Gibbs sampler for $\lambda_{i,t}^{(m)}$, from $\lambda_{i,t}^{(m)} \sim LN(e^{\mu_{i,t}+\frac{\sigma^2}{2}}, (e^{\mu_{i,t}} - 1)e^{2\mu_{i,t}+\sigma^2})$, and accept it as the mth draw with probability

$$\alpha\left(\lambda_{i,t}^{(m-1)}, \lambda_{i,t}^{(m)}\right) = \min\{1, \frac{f\left(y_{i,t}^{*} \mid \lambda_{i,t}^{(m)}\right)}{f\left(y_{i,t}^{*} \mid \lambda_{i,t}^{(m-1)}\right)}\}, \quad (70)$$

since the proposal densities cancel out in the ratio.

5. Draw from the posterior density of the diagonal elements of $\mathbf{\Phi}$ conditional on $\mathbf{\Pi}$, $\mathbf{\Psi}$, $\mathbf{B}$, and $\underline{\mathbf{\Lambda_T}}$ and the data.

   The Inverse Gamma prior is conjugate for the variance parameter of the Normal

density. Therefore, the conditional posterior of φ_i is also Inverse Gamma as

$$f(\varphi_i|\lambda_i) \propto h\left(\lambda_i \mid \varphi_i\right) p(\varphi_i)$$
$$\propto \prod_{t=1}^{T} \frac{1}{\varphi_i^{0.5}} \exp\left\{-\frac{\left(ln\left(\lambda_{i,t}\right) - ln\left(\lambda_{i,t-1}\right)\right)^2}{2\varphi_i}\right\} \times \varphi_i^{-\left(\frac{\gamma}{2}+1\right)} e^{-\frac{\delta}{2\varphi_i}}. \quad (71)$$

Furthermore, the right hand side above is equal to

$$\varphi_i^{-\left(\frac{\gamma}{2}+1\right)-\frac{T}{2}} \exp\left\{-\frac{\delta}{2\varphi_i} - \frac{1}{2\varphi_i}\sum_{t=1}^{T} ln\left(\frac{\lambda_{i,t}}{\lambda_{i,t-1}}\right)^2\right\}$$
$$= \varphi_i^{-\left(\frac{\gamma+T}{2}+1\right)} \exp\left\{-\frac{\delta + \sum_{t=1}^{T} ln\left(\frac{\lambda_{i,t}}{\lambda_{i,t-1}}\right)^2}{2\varphi_i}\right\} \quad (72)$$

Consequently, assuming identical Inverse Gamma priors on each $\varphi_i \sim IG(\frac{\gamma}{2}, \frac{\delta}{2})$, the conditional posterior is also Inverse Gamma, or $IG(\frac{\gamma^*}{2}, \frac{\delta^*}{2})$, where

$$\gamma^* = \gamma + \mathrm{T}, \quad (73a)$$

$$\text{and} \quad \delta^* = \delta + \sum_{t=1}^{T} \left(\ln\left(\frac{\lambda_{i,t}}{\lambda_{i,t-1}}\right)\right)^2. \quad (73b)$$

11.3 Computation of the posterior distribution for the IWSV model

Let us now describe the sequence of conditional posterior distributions for the $IWSV$ model.

1. First, we repeat step (1) above in Section 11.2, except that we replace $\mathbf{\Gamma_t}=\mathbf{B}^{-1}\mathbf{\Lambda_t}(\mathbf{B}^{-1})' = var\left(\mathbf{v_t}\right)$ with $\mathbf{\Sigma_t}$. That is, we no longer condition on $\mathbf{B}$, $\underline{\mathbf{\Lambda_T}}$, and $\mathbf{\Phi}$, but rather on $\mathbf{A_k}$, for $k = 1, \ldots, K$, $\mathbf{C}$, ν, and $\underline{\mathbf{\Sigma_T}}$.

2. Repeat step (2) above in Section 11.2, except this time replace $\mathbf{\Gamma_t}=\mathbf{B}^{-1}\mathbf{\Lambda_t}(\mathbf{B}^{-1})' = var\left(\mathbf{v_t}\right)$, with $\mathbf{\Sigma_t}$.

3. Draw from the posterior density of the parameters $\mathbf{A_k},\ \forall k, \mathbf{C}$, and ν jointly, conditional on $\mathbf{\Pi}$, $\mathbf{\Psi}$, and $\underline{\mathbf{\Sigma_T}}$, and the data.

   All of the individual elements of the parameter matrices $\mathbf{A_k},\ \forall k, \mathbf{C}$ and ν are drawn jointly by a Metropolis-within-Gibbs step employing a random walk proposal. The joint proposal is multivariate Normal, and we assume multivariate Normal priors on both $\mathbf{A_k},\ \forall k$ and $\mathbf{C}$ and a Gamma prior on $(\nu - p)$. See Section 4 on priors for more details.

   The random walk multivariate Normal proposal is symmetric and conditioned on the last value in the process through its mean vector; therefore it drops out of the acceptance ratio. The variance of the proposal is initially set to the inverse of the observed negative Hessian matrix at the mode of the conditional posterior for a first attempt, and then a second attempt is employed using the covariance matrix of the initial Markov process draws themselves for improved mixing.

   Moreover, the likelihood of the $IWSV$ model is now given as

$$\begin{aligned} f\left(\mathbf{v} \mid \theta\right) = L\left(\theta\right) &= \prod_{t=1}^{\mathrm{T}} f\left(\mathbf{v_t} \mid \mathbf{\Sigma_t}\right) g\left(\mathbf{\Sigma_t} \mid \mathbf{\Sigma_{t-1}}, \ldots, \mathbf{\Sigma_{t-K}}; \theta\right) \\ &= \prod_{t=1}^{T} \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{\Sigma_t}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\mathbf{v_t}^{'}\mathbf{\Sigma_t}^{-1}\mathbf{v_t}\right\} \\ \times 2^{-\left(\frac{\nu p}{2}\right)} |\mathbf{S_{t-1}}|^{\frac{\nu}{2}} \Gamma_p\left(\frac{\nu}{2}\right)^{-1} &|\mathbf{\Sigma_t}|^{-(\nu+p+1)/2} \exp\left\{-\frac{1}{2} tr\left[\mathbf{S_{t-1}}\mathbf{\Sigma_t}^{-1}\right]\right\}, \end{aligned} \quad (74)$$

   where $\mathbf{v_t} = \mathbf{\Pi}\left(L\right)\left(\mathbf{y_t} - \mathbf{\Psi d_t}\right)$ is a function of the data, $\mathbf{y}$. Therefore, by Bayes Theorem we can consider the conditional posterior of θ as proportional to the likelihood (which is really a function of the data) times the prior density for θ (where $\theta = \{\mathbf{A_1}, \ldots, \mathbf{A_k}, \mathbf{C}, \nu\}$) as follows

$$p(\theta|\underline{\mathbf{y_T}}, \mathbf{\Pi}, \mathbf{\Psi}, \underline{\mathbf{\Sigma_T}}) \propto L(\theta)\pi(\theta) = f(\underline{\mathbf{y_T}}, \underline{\mathbf{\Sigma_T}}|\theta; \mathbf{\Pi}, \mathbf{\Psi})\pi(\theta) \propto f(\underline{\mathbf{v_T}}|\theta)\pi(\theta). \quad (75)$$

   Therefore, the Metropolis acceptance probability of the mth draw, $\theta^{(m)}$, in the

random walk sampler can be expressed as

$$\alpha(\theta^{(m-1)}, \theta^{(m)}) = \min\left\{1, \frac{p(\theta^{(m)}|\underline{\mathbf{y_T}}, \mathbf{\Pi}, \mathbf{\Psi}, \underline{\mathbf{\Sigma_T}})}{p(\theta^{(m-1)}|\underline{\mathbf{y_T}}, \mathbf{\Pi}, \mathbf{\Psi}, \underline{\mathbf{\Sigma_T}})}\right\}. \tag{76}$$

4. Similarly to step (4) above, we now draw from the posterior density of $\mathbf{\Sigma_t}$ conditional on
$\mathbf{\Sigma_{\backslash t}}, \mathbf{A_k}, \forall k, \mathbf{C}, \nu, \mathbf{\Pi}, \mathbf{\Psi}$, and the data, in sequence for each time $t = 1, \ldots, T$. We have:

$$P\left(\mathbf{\Sigma_t} \mid \mathbf{\Sigma_{\backslash t}}, \mathbf{v}\right) \propto P\left(\mathbf{v_t} \mid \mathbf{\Sigma_t}\right) P\left(\mathbf{\Sigma_t} \mid \mathbf{\Sigma_{t-1}}\right) P\left(\mathbf{\Sigma_{t+1}} \mid \mathbf{\Sigma_t}\right) \tag{77a}$$

$$\propto |\mathbf{\Sigma_t}|^{-\frac{1}{2}} |\mathbf{S_t}|^{\frac{\nu}{2}} |\mathbf{\Sigma_t}|^{-(\nu+p+1)/2} \exp\left\{-\frac{1}{2} tr\left[\left(\mathbf{S_{t-1}} + \mathbf{v_t v_t'}\right)\mathbf{\Sigma_t^{-1}}\right]\right\}$$

$$\exp\left\{-\frac{1}{2} tr\left[\mathbf{S_t \Sigma_{t+1}^{-1}}\right]\right\}, \tag{77b}$$

$$\text{where} \quad \frac{\mathbf{S_{t-1}}}{\nu - p - 1} = \mathbf{CC'} + \sum_{k=1}^{K} \mathbf{A}_k \mathbf{\Sigma_{t-k}^{-1}} \mathbf{A}_k'. \tag{77c}$$

Therefore, by letting the proposal be Inverse Wishart $\mathbf{\Sigma_t} \sim IW_p\left(\nu, \mathbf{S_{t-1}^*}\right)$ where $\mathbf{S_{t-1}^*} = \mathbf{S_{t-1}} + \mathbf{v_t v_t'}$, the proposal drops out of the Metropolis-Hastings ratio. Indeed, the probability of accepting the mth draw of $\mathbf{\Sigma_t^{(m)}}$, sequentially, for each time period $t = 1, \ldots, T$, is now [19]

$$\alpha\left(\mathbf{\Sigma_t^{(m-1)}}, \mathbf{\Sigma_t^{(m)}}\right) = \min\left\{1, \frac{\left|\mathbf{\Sigma_t^{(m)}}\right|^{-\frac{1}{2}} \left|\mathbf{S_t^{(m)}}\right|^{\frac{\nu}{2}} \exp\left\{-\frac{1}{2} tr\left[\mathbf{S_t^{(m)}} \mathbf{\Sigma_{t+1}^{-1}}\right]\right\}}{\left|\mathbf{\Sigma_t^{(m-1)}}\right|^{-\frac{1}{2}} \left|\mathbf{S_t^{(m-1)}}\right|^{\frac{\nu}{2}} \exp\left\{-\frac{1}{2} tr\left[\mathbf{S_t^{(m-1)}} \mathbf{\Sigma_{t+1}^{-1}}\right]\right\}}\right\}. \tag{78}$$

12 Appendix: Tables and Figures

[19] To avoid numerical problems logs are taken of both the numerator and denominator, then differenced, before finally taking their exponential. This avoids issues when the non-logged function values grow either too large or too small to be machine comparable.

Table 3.i: Section 7.1: Posterior distribution of the parameters for $IWSV$ model, Simulated data, 1st sample, Rolling window

Parameter	Pop. value	2.5% C.I.	mean	97.5% C.I.	Parameter	Pop. value	2.5% C.I.	mean	97.5% C.I.
$vech(\mathbf{C})$	0.3	0.2778	0.2984	0.3208	$vec(\mathbf{\Pi_1})$	0.2	0.0976	0.2185	0.3374
	0	-0.0229	0.0046	0.0320		0	-0.1666	-0.0157	0.1360
	0	-0.0413	-0.0069	0.0248		0	-0.2724	-0.0920	0.0882
	0	-0.0407	-0.0048	0.0381		0	-0.2415	0.0110	0.2627
	0.3	0.2775	0.3038	0.3337		0	-0.0806	0.0113	0.1025
	0	-0.0375	-0.0057	0.0273		0.8	0.5904	0.7091	0.8277
	0	-0.0392	-0.0003	0.0418		0	-0.0888	0.0456	0.1798
	0.3	0.2736	0.3077	0.3427		0	-0.1756	0.0218	0.2188
	0	-0.0468	-0.0039	0.0403		0	-0.1156	-0.0425	0.0309
	0.3	0.2701	0.3138	0.3577		0	-0.0961	-0.0047	0.0870
$vec(\mathbf{A_1})$	0.5	0.4449	0.4943	0.5344		0.8	0.7345	0.8532	0.9721
	0	-0.0417	0.0020	0.0486		0	-0.2937	-0.1246	0.0447
	0	-0.0438	0.0019	0.0455		0	-0.0508	-0.0107	0.0291
	0	-0.0468	0.0014	0.0503		0	0.0227	0.0744	0.1262
	0	-0.0467	-0.0018	0.0487		0	-0.1318	-0.0693	-0.0065
	0.75	0.7062	0.7440	0.7811		0.8	0.6551	0.7704	0.8863
	0	-0.0477	-0.0029	0.0421	$vec(\mathbf{\Pi_2})$	0	-0.2741	-0.1557	-0.0367
	0	-0.0466	-0.0011	0.0451		0	-0.2246	-0.0768	0.0697
	0	-0.0447	-0.0077	0.0338		0	-0.1544	0.0256	0.2042
	0	-0.0444	-0.0030	0.0349		0	-0.2398	0.0054	0.2486
	0.85	0.8052	0.8459	0.8804		0	-0.1330	-0.0199	0.0927
	0	-0.0417	-0.0008	0.0415		0	-0.0688	0.0753	0.2175
	0	-0.0200	0.0027	0.0293		0	-0.2405	-0.0768	0.0860
	0	-0.0191	0.0039	0.0267		0	-0.1928	0.0324	0.2563
	0	-0.0250	0.0024	0.0279		0	-0.0284	0.0688	0.1653
	0.98	0.9477	0.9739	1.0006		0	-0.1910	-0.0714	0.0485
$vec(\mathbf{A_2})$	0	0.0007	0.0189	0.0508		0	-0.1478	0.0007	0.1505
	0	-0.0455	-0.0005	0.0402		0	-0.1146	0.0918	0.2959
	0	-0.0424	0.0001	0.0421		0	-0.0796	-0.0276	0.0244
	0	-0.0445	0.0016	0.0456		0	-0.0540	0.0124	0.0781
	0	-0.0416	0.0002	0.0460		0	0.0106	0.0902	0.1691
	0	-0.0438	-0.0014	0.0400		0	-0.0797	0.0618	0.2005
	0	-0.0458	-0.0024	0.0458	$vec(\mathbf{\Pi_3})$	0	-0.1753	-0.0536	0.0676
	0	-0.0440	-0.0007	0.0428		0	-0.0817	0.0704	0.2227
	0	-0.0452	-0.0003	0.0453		0	-0.0947	0.0814	0.2565
	0	-0.0409	0.0010	0.0412		0	-0.3865	-0.1355	0.1140
	0	-0.0487	0.0017	0.0479		0	-0.0670	0.0241	0.1144
	0	-0.0488	-0.0033	0.0466		0	-0.1501	-0.0357	0.0786
	0	-0.0372	0.0016	0.0433		0	-0.1332	-0.0015	0.1319
	0	-0.0421	-0.0007	0.0410		0	-0.1896	0.0080	0.2093
	0	-0.0450	-0.0002	0.0431		0	-0.1138	-0.0421	0.0296
	0	-0.0381	0.0018	0.0450		0	-0.0385	0.0546	0.1461
$vec(\mathbf{A_3})$	0	0.0005	0.0158	0.0448		0	-0.1840	-0.0714	0.0401
	0	-0.0382	0.0021	0.0477		0	-0.2107	-0.0451	0.1206
	0	-0.0470	-0.0011	0.0464		0	-0.0360	0.0064	0.0488
	0	-0.0499	-0.0001	0.0445		0	-0.1245	-0.0708	-0.0177
	0	-0.0428	0.0005	0.0423		0	-0.0620	0.0039	0.0705
	0	-0.0422	0.0001	0.0404		0	-0.1837	-0.0698	0.0439
	0	-0.0434	-0.0046	0.0376	$\mathbf{\Psi}$	3	2.9699	3.0083	3.0465
	0	-0.0420	0.0006	0.0432		0	-0.0832	-0.0223	0.0388
	0	-0.0478	-0.0032	0.0379		2.5	2.4380	2.4987	2.5600
	0	-0.0452	-0.0037	0.0425		0	-0.0596	0.0016	0.0630
	0	-0.0459	-0.0002	0.0483					
	0	-0.0424	-0.0003	0.0441					
	0	-0.0387	-0.0001	0.0388					
	0	-0.0396	0.0022	0.0399					
	0	-0.0396	0.0040	0.0544					
	0	-0.0395	0.0041	0.0477					
ν	30	15.9899	19.9123	26.1828					

Table 3.ii: Section 7.1: Distribution of the posterior mean of the parameters for $IWSV$ model, Simulated data, across $N = 100$ samples, Rolling window

Parameter	Pop. value	2.5% C.I.	mean	97.5% C.I.	Parameter	Pop. value	2.5% C.I.	mean	97.5% C.I.
$vech(\mathbf{C})$	0.3	0.2943	0.3003	0.3070	$vec(\mathbf{\Pi_1})$	0.2	0.2145	0.2303	0.2498
	0	-0.0003	0.0033	0.0086		0	-0.0441	-0.0107	0.0206
	0	-0.0146	-0.0093	-0.0053		0	-0.1211	-0.0611	0.0006
	0	-0.0105	-0.0051	0.0005		0	-0.0394	-0.0132	0.0140
	0.3	0.3032	0.3098	0.3155		0	0.0114	0.0360	0.0522
	0	-0.0133	-0.0068	-0.0001		0.8	0.6974	0.7248	0.7497
	0	-0.0069	-0.0019	0.0031		0	-0.0833	-0.0031	0.0751
	0.3	0.2924	0.2994	0.3064		0	-0.0148	0.0361	0.0751
	0	-0.0041	0.0007	0.0059		0	-0.0377	-0.0055	0.0193
	0.3	0.3101	0.3147	0.3192		0	-0.0301	-0.0124	-0.0005
$vec(\mathbf{A_1})$	0.5	0.4921	0.4970	0.5031		0.8	0.6782	0.7691	0.8529
	0	-0.0052	-0.0004	0.0048		0	-0.1299	-0.1068	-0.0662
	0	-0.0050	-0.0011	0.0028		0	-0.0074	-0.0009	0.0056
	0	-0.0069	-0.0016	0.0029		0	0.0089	0.0418	0.0747
	0	-0.0022	0.0038	0.0106		0	-0.0901	-0.0726	-0.0578
	0.75	0.7377	0.7438	0.7492		0.8	0.7667	0.7887	0.8148
	0	-0.0060	-0.0005	0.0046	$vec(\mathbf{\Pi_2})$	0	-0.1492	-0.1286	-0.1064
	0	-0.0098	-0.0041	0.0019		0	-0.0638	-0.0408	-0.0230
	0	-0.0127	-0.0053	0.0031		0	-0.0440	0.0070	0.0423
	0	-0.0105	-0.0024	0.0052		0	0.0107	0.0671	0.1271
	0.85	0.8334	0.8408	0.8481		0	-0.0437	-0.0253	-0.0119
	0	-0.0058	-0.0014	0.0034		0	0.0112	0.0404	0.0696
	0	-0.0134	-0.0056	0.0040		0	-0.1043	-0.0603	0.0024
	0	0.0000	0.0051	0.0120		0	-0.0987	-0.0476	0.0028
	0	-0.0043	0.0018	0.0088		0	0.0368	0.0551	0.0869
	0.98	0.9675	0.9743	0.9825		0	-0.0745	-0.0430	-0.0105
$vec(\mathbf{A_2})$	0	0.0156	0.0178	0.0208		0	-0.0068	0.0230	0.0750
	0	-0.0058	0.0001	0.0050		0	-0.0123	0.0596	0.1291
	0	-0.0044	0.0000	0.0044		0	-0.0298	-0.0116	0.0031
	0	-0.0052	0.0001	0.0051		0	-0.0047	0.0103	0.0240
	0	-0.0037	0.0000	0.0043		0	0.0596	0.0895	0.1091
	0	-0.0055	-0.0003	0.0047		0	0.0423	0.0695	0.1054
	0	-0.0039	0.0003	0.0051	$vec(\mathbf{\Pi_3})$	0	-0.0437	-0.0268	-0.0100
	0	-0.0042	0.0001	0.0049		0	-0.0723	-0.0117	0.0631
	0	-0.0043	-0.0001	0.0041		0	0.0212	0.0502	0.0785
	0	-0.0049	0.0004	0.0043		0	-0.1623	-0.1254	-0.0914
	0	-0.0055	-0.0002	0.0047		0	0.0073	0.0316	0.0690
	0	-0.0053	0.0001	0.0057		0	-0.0352	-0.0194	-0.0017
	0	-0.0044	0.0000	0.0037		0	-0.0057	0.0414	0.0755
	0	-0.0047	-0.0003	0.0043		0	-0.0033	0.0635	0.1190
	0	-0.0040	0.0006	0.0052		0	-0.0634	-0.0495	-0.0390
	0	-0.0042	0.0001	0.0052		0	0.0088	0.0386	0.0664
$vec(\mathbf{A_3})$	0	0.0147	0.0178	0.0206		0	-0.0939	-0.0620	-0.0218
	0	-0.0045	0.0000	0.0050		0	-0.1322	-0.0835	-0.0406
	0	-0.0048	0.0001	0.0047		0	-0.0266	-0.0082	0.0124
	0	-0.0045	0.0003	0.0053		0	-0.0705	-0.0422	-0.0259
	0	-0.0044	0.0000	0.0054		0	-0.0268	-0.0144	0.0048
	0	-0.0042	0.0001	0.0042		0	-0.0778	-0.0592	-0.0424
	0	-0.0057	0.0000	0.0049	$\mathbf{\Psi}$	3	3.0092	3.0289	3.0482
	0	-0.0043	0.0000	0.0046		0	-0.0236	-0.0207	-0.0163
	0	-0.0059	0.0002	0.0054		2.5	2.4865	2.4942	2.5011
	0	-0.0040	0.0002	0.0052		0	0.0001	0.0015	0.0027
	0	-0.0051	0.0002	0.0052					
	0	-0.0031	0.0002	0.0036					
	0	-0.0038	0.0001	0.0038					
	0	-0.0045	0.0000	0.0048					
	0	-0.0042	0.0002	0.0048					
	0	-0.0057	-0.0002	0.0050					
ν	30	18.2853	19.8090	21.3300					

Table 4.i: Section 7.1: Posterior distribution of the parameters for $Clark$ model, Simulated data, 1st sample, Rolling window

Parameter	Pop. value	2.5% C.I.	mean	97.5% C.I.	Parameter	Pop. value	2.5% C.I.	mean	97.5% C.I.
b_{21}	n/a	-0.1449	0.0192	0.1800	$vec(\mathbf{\Pi_1})$	0.2	0.0839	0.2117	0.3408
b_{31}		-0.0650	0.1480	0.3667		0	-0.1787	-0.0310	0.1214
b_{32}		-0.0766	0.0693	0.2143		0	-0.2839	-0.0989	0.0873
b_{41}		-0.3200	0.0221	0.3661		0	-0.2180	0.0422	0.3032
b_{42}		-0.2173	0.0609	0.3390		0	-0.0723	0.0170	0.1059
b_{43}		-0.3054	-0.0779	0.1498		0.8	0.5884	0.7131	0.8391
φ_1		0.1643	0.2869	0.4712		0	-0.1150	0.0198	0.1573
φ_2		0.1624	0.2783	0.4575		0	-0.2074	-0.0027	0.2066
φ_3		0.1625	0.2784	0.4718		0	-0.1155	-0.0420	0.0307
φ_4		0.1655	0.2857	0.4731		0	-0.1196	-0.0223	0.0741
						0.8	0.7102	0.8354	0.9611
						0	-0.3274	-0.1493	0.0291
						0	-0.0555	-0.0175	0.0211
						0	0.0135	0.0681	0.1237
						0	-0.1441	-0.0806	-0.0168
						0.8	0.6478	0.7690	0.8883
					$vec(\mathbf{\Pi_2})$	0	-0.2772	-0.1491	-0.0207
						0	-0.2267	-0.0713	0.0821
						0	-0.1454	0.0432	0.2280
						0	-0.2279	0.0210	0.2680
						0	-0.1414	-0.0332	0.0764
						0	-0.0729	0.0736	0.2188
						0	-0.2103	-0.0430	0.1203
						0	-0.1926	0.0414	0.2765
						0	-0.0142	0.0742	0.1621
						0	-0.1774	-0.0584	0.0619
						0	-0.1553	-0.0034	0.1481
						0	-0.0863	0.1282	0.3420
						0	-0.0602	-0.0109	0.0385
						0	-0.0536	0.0157	0.0842
						0	0.0181	0.0985	0.1791
						0	-0.0803	0.0655	0.2081
					$vec(\mathbf{\Pi_3})$	0	-0.1198	0.0034	0.1265
						0	-0.0921	0.0618	0.2147
						0	-0.1051	0.0763	0.2574
						0	-0.3904	-0.1337	0.1235
						0	-0.0537	0.0316	0.1162
						0	-0.1479	-0.0305	0.0875
						0	-0.1297	0.0010	0.1327
						0	-0.2270	-0.0112	0.2051
						0	-0.1037	-0.0370	0.0295
						0	-0.0458	0.0465	0.1379
						0	-0.1650	-0.0478	0.0695
						0	-0.2126	-0.0398	0.1322
						0	-0.0392	0.0020	0.0431
						0	-0.1223	-0.0666	-0.0128
						0	-0.0792	-0.0091	0.0607
						0	-0.1899	-0.0723	0.0474
					$\mathbf{\Psi}$	3	2.9584	2.9990	3.0399
						0	-0.0846	-0.0235	0.0378
						2.5	2.4360	2.4970	2.5587
						0	-0.0591	0.0024	0.0637

Table 4.ii: Section 7.1: Distribution of the posterior mean of the parameters for $Clark$ model, Simulated data, across $N = 100$ samples, Rolling window

Parameter	Pop. value	2.5% C.I.	mean	97.5% C.I.	Parameter	Pop. value	2.5% C.I.	mean	97.5% C.I.
b_{21}	n/a	-0.0622	-0.0161	0.0575	$vec(\mathbf{\Pi_1})$	0.2	0.2074	0.2277	0.2505
b_{31}		0.0581	0.1111	0.1570		0	-0.0475	-0.0157	0.0198
b_{32}		0.0447	0.0737	0.0976		0	-0.1249	-0.0656	-0.0187
b_{41}		0.0136	0.1623	0.2286		0	-0.0248	0.0012	0.0265
b_{42}		-0.0571	-0.0100	0.0425		0	0.0155	0.0661	0.1033
b_{43}		-0.1218	-0.0750	0.0001		0.8	0.6895	0.7079	0.7328
φ_1		0.2860	0.2984	0.3196		0	-0.0761	-0.0090	0.0645
φ_2		0.2774	0.2914	0.3149		0	-0.0254	0.0169	0.0461
φ_3		0.2659	0.2872	0.3172		0	-0.0409	-0.0096	0.0230
φ_4		0.2744	0.2997	0.3161		0	-0.0643	-0.0407	-0.0207
						0.8	0.6899	0.7621	0.8339
						0	-0.1470	-0.1236	-0.0946
						0	-0.0179	-0.0070	0.0022
						0	0.0006	0.0443	0.0703
						0	-0.1090	-0.0846	-0.0703
						0.8	0.7640	0.7833	0.8128
					$vec(\mathbf{\Pi_2})$	0	-0.1434	-0.1120	-0.0737
						0	-0.0686	-0.0317	-0.0004
						0	0.0121	0.0386	0.0704
						0	0.0286	0.0782	0.1441
						0	-0.0708	-0.0528	-0.0257
						0	0.0084	0.0390	0.0689
						0	-0.0699	-0.0414	0.0119
						0	-0.0815	-0.0274	0.0187
						0	0.0504	0.0684	0.0995
						0	-0.0635	-0.0325	0.0010
						0	-0.0191	0.0215	0.0754
						0	0.0013	0.0799	0.1460
						0	-0.0150	0.0037	0.0214
						0	-0.0108	0.0104	0.0262
						0	0.0854	0.1053	0.1303
						0	0.0481	0.0778	0.1082
					$vec(\mathbf{\Pi_3})$	0	-0.0087	0.0169	0.0382
						0	-0.0770	-0.0212	0.0484
						0	0.0023	0.0451	0.0832
						0	-0.1576	-0.1200	-0.0886
						0	0.0166	0.0399	0.0680
						0	-0.0290	-0.0019	0.0201
						0	-0.0174	0.0305	0.0564
						0	-0.0172	0.0443	0.0981
						0	-0.0677	-0.0514	-0.0359
						0	0.0155	0.0505	0.0738
						0	-0.0816	-0.0649	-0.0520
						0	-0.1280	-0.0889	-0.0445
						0	-0.0359	-0.0135	0.0114
						0	-0.0660	-0.0394	-0.0201
						0	-0.0458	-0.0301	-0.0060
						0	-0.0893	-0.0699	-0.0503
					$\mathbf{\Psi}$	3	2.9994	3.0256	3.0471
						0	-0.0282	-0.0236	-0.0181
						2.5	2.4833	2.4926	2.5009
						0	-0.0003	0.0019	0.0034

Table 5.i: Section 7.2: Posterior distribution of the parameters for $IWSV$ model, Real data, 1st sample, Recursive window

Parameter	Prior mean	2.5% C.I.	Posterior mean	97.5% C.I.	Parameter	Prior mean	2.5% C.I.	Posterior mean	97.5% C.I.
$vech(\mathbf{C})$	0.3	0.3026	0.3668	0.4314	$vec(\mathbf{\Pi_1})$	0.2	0.0998	0.2678	0.4365
	0	-0.0924	-0.0101	0.0665		0	-0.0432	0.0129	0.0696
	0	-0.0625	0.0089	0.0712		0	-0.0086	0.0187	0.0462
	0	-0.0526	0.0009	0.0526		0	-0.0456	-0.0288	-0.0128
	0.3	0.3004	0.3702	0.4730		0	0.2734	0.6964	1.1389
	0	-0.0661	-0.0065	0.0591		0.8	0.3613	0.5683	0.7760
	0	-0.0706	-0.0230	0.0256		0	-0.0551	0.0092	0.0750
	0.3	0.1957	0.2553	0.3155		0	-0.0602	-0.0193	0.0190
	0	-0.0541	-0.0050	0.0480		0	-0.6476	-0.1205	0.4230
	0.3	0.0882	0.1251	0.1671		0	-0.0750	0.1870	0.4571
$vec(\mathbf{A_1})$	0.9	0.9828	1.0039	1.0259		0.8	1.0069	1.1873	1.3663
	0	-0.0151	0.0101	0.0246		0	-0.0583	0.0105	0.0761
	0	-0.0046	0.0046	0.0140		0	-0.6838	0.0451	0.7751
	0	-0.0163	-0.0098	-0.0039		0	-0.7457	-0.2081	0.3390
	0	-0.1494	-0.0655	0.0745		0	-0.3954	-0.0664	0.2640
	0.9	0.8988	0.9514	0.9860		0.8	0.9701	1.1614	1.3534
	0	-0.0364	-0.0029	0.0260	$vec(\mathbf{\Pi_2})$	0	-0.0961	0.0635	0.2214
	0	-0.0089	0.0083	0.0255		0	-0.0348	0.0229	0.0795
	0	-0.0757	0.0047	0.0844		0	-0.0180	0.0107	0.0391
	0	-0.0931	-0.0094	0.0740		0	-0.0378	-0.0224	-0.0072
	0.9	0.8054	0.8714	0.9351		0	-0.9829	-0.5763	-0.1504
	0	-0.0614	-0.0155	0.0244		0	-0.0972	0.0953	0.2925
	0	-0.1045	-0.0133	0.0735		0	-0.0895	-0.0237	0.0432
	0	-0.1191	-0.0446	0.0536		0	-0.0280	0.0114	0.0504
	0	-0.0744	0.0077	0.0824		0	-0.9201	-0.2808	0.3699
	0.9	0.7397	0.8091	0.8692		0	-0.4920	-0.1107	0.2704
$vec(\mathbf{A_2})$	0	0.0017	0.0405	0.1098		0	-0.7450	-0.4738	-0.2042
	0	-0.0613	-0.0032	0.0616		0	-0.1357	-0.0376	0.0629
	0	-0.0280	0.0015	0.0281		0	-0.3902	0.3363	1.0635
	0	-0.0170	-0.0016	0.0141		0	-0.4375	0.1732	0.7828
	0	-0.0974	-0.0062	0.0935		0	-0.3910	0.0464	0.4760
	0	-0.0970	-0.0014	0.0931		0	-0.6903	-0.4303	-0.1641
	0	-0.0621	0.0035	0.0640	$vec(\mathbf{\Pi_3})$	0	-0.3406	-0.1860	-0.0356
	0	-0.0439	0.0000	0.0385		0	-0.0329	0.0165	0.0665
	0	-0.1015	-0.0019	0.0791		0	-0.0300	-0.0042	0.0214
	0	-0.0829	-0.0003	0.0934		0	-0.0161	-0.0025	0.0115
	0	-0.0854	-0.0009	0.0903		0	-0.6159	-0.2589	0.0889
	0	-0.0677	-0.0023	0.0659		0	-0.1229	0.0435	0.2100
	0	-0.1018	-0.0074	0.0855		0	-0.0574	-0.0055	0.0477
	0	-0.1016	-0.0071	0.0795		0	-0.0161	0.0169	0.0490
	0	-0.0804	0.0087	0.0984		0	-0.3772	0.2538	0.8965
	0	-0.0952	-0.0038	0.0792		0	-0.2421	0.0972	0.4394
$vec(\mathbf{A_3})$	0	0.0022	0.0427	0.1221		0	0.0530	0.2489	0.4410
	0	-0.0603	0.0011	0.0560		0	-0.0582	0.0239	0.1064
	0	-0.0282	0.0011	0.0300		0	-0.4226	0.2740	0.9774
	0	-0.0144	0.0012	0.0171		0	-0.4789	0.0136	0.5048
	0	-0.0911	-0.0016	0.0869		0	-0.3233	-0.0316	0.2655
	0	-0.0893	0.0014	0.0916		0	-0.0173	0.1529	0.3212
	0	-0.0596	-0.0047	0.0552	$\mathbf{\Psi}$	3	2.9048	3.2533	3.5801
	0	-0.0403	0.0043	0.0457		0	-0.1817	0.1979	0.5711
	0	-0.0712	0.0140	0.1038		2.5	0.8292	1.8164	3.0835
	0	-0.0958	-0.0006	0.0941		0	-0.1466	0.2088	0.5409
	0	-0.0922	-0.0082	0.0807					
	0	-0.0632	0.0059	0.0777					
	0	-0.0802	0.0066	0.0923					
	0	-0.0924	0.0057	0.0911					
	0	-0.0755	0.0048	0.0895					
	0	-0.0822	0.0020	0.0853					
ν	15	13.0737	15.9185	19.2311					

Table 5.ii: Section 7.2: Distribution of the posterior mean of the parameters for $IWSV$ model, Real data, across $N = 100$ samples, Recursive window

Parameter	Prior mean	2.5% C.I.	Distribution mean	97.5% C.I.	Parameter	Prior mean	2.5% C.I.	Distribution mean	97.5% C.I.
$vech(\mathbf{C})$	0.3	0.3169	0.3570	0.3940	$vec(\mathbf{\Pi_1})$	0.2	0.1783	0.2069	0.2417
	0	-0.0472	-0.0101	0.0244		0	0.0188	0.0283	0.0401
	0	-0.0102	0.0087	0.0266		0	0.0138	0.0169	0.0203
	0	-0.0098	0.0029	0.0156		0	-0.0314	-0.0254	-0.0222
	0.3	0.3078	0.3575	0.3972		0	0.2150	0.3770	0.6117
	0	-0.0137	-0.0019	0.0135		0.8	0.5636	0.5971	0.6360
	0	-0.0271	-0.0174	-0.0100		0	0.0000	0.0181	0.0254
	0.3	0.1809	0.2132	0.2406		0	-0.0215	-0.0054	0.0013
	0	-0.0202	-0.0078	0.0042		0	-0.0235	0.1114	0.2094
	0.3	0.0946	0.1171	0.1378		0	0.0029	0.0524	0.1450
$vec(\mathbf{A_1})$	0.9	0.9962	1.0018	1.0068		0.8	1.1440	1.2516	1.3223
	0	-0.0110	0.0013	0.0104		0	-0.0072	-0.0019	0.0104
	0	0.0023	0.0047	0.0076		0	-0.0303	0.0498	0.1240
	0	-0.0112	-0.0100	-0.0089		0	-0.2833	-0.2083	-0.1313
	0	-0.0792	-0.0297	0.0369		0	-0.1228	-0.0924	-0.0575
	0.9	0.9218	0.9332	0.9467		0.8	1.1078	1.1842	1.2344
	0	-0.0033	0.0036	0.0091	$vec(\mathbf{\Pi_2})$	0	0.0807	0.1457	0.1885
	0	-0.0004	0.0040	0.0076		0	0.0078	0.0137	0.0208
	0	-0.0436	-0.0006	0.0357		0	0.0071	0.0107	0.0155
	0	-0.0238	0.0082	0.0414		0	-0.0234	-0.0186	-0.0150
	0.9	0.8701	0.8866	0.8998		0	-0.5158	-0.3684	-0.2794
	0	-0.0247	-0.0191	-0.0134		0	0.0854	0.1547	0.1928
	0	-0.0335	-0.0008	0.0359		0	-0.0230	-0.0055	0.0048
	0	-0.0429	-0.0184	0.0141		0	0.0070	0.0103	0.0155
	0	-0.0244	-0.0019	0.0221		0	-0.4067	-0.2102	-0.0803
	0.9	0.7822	0.7959	0.8100		0	-0.1261	0.0442	0.1253
$vec(\mathbf{A_2})$	0	0.0363	0.0452	0.0544		0	-0.5686	-0.5257	-0.4351
	0	-0.0141	-0.0011	0.0110		0	-0.0389	-0.0268	-0.0187
	0	-0.0026	0.0011	0.0056		0	0.1418	0.3322	0.4701
	0	-0.0029	-0.0014	0.0005		0	0.0574	0.1723	0.2386
	0	-0.0115	-0.0004	0.0103		0	-0.0120	0.0219	0.0637
	0	-0.0182	-0.0062	0.0060		0	-0.4234	-0.3567	-0.3030
	0	-0.0085	-0.0007	0.0072	$vec(\mathbf{\Pi_3})$	0	-0.1889	-0.1567	-0.1048
	0	-0.0021	0.0019	0.0054		0	-0.0003	0.0141	0.0229
	0	-0.0127	0.0011	0.0144		0	-0.0102	-0.0065	-0.0027
	0	-0.0109	0.0021	0.0172		0	-0.0061	0.0004	0.0043
	0	-0.0202	-0.0054	0.0067		0	-0.3499	-0.2877	-0.2174
	0	-0.0081	-0.0001	0.0064		0	0.0297	0.1074	0.1437
	0	-0.0120	-0.0005	0.0097		0	-0.0171	-0.0132	-0.0063
	0	-0.0109	0.0002	0.0121		0	0.0170	0.0217	0.0255
	0	-0.0084	0.0005	0.0120		0	-0.0871	0.0283	0.1934
	0	-0.0108	0.0000	0.0080		0	-0.1700	-0.0872	0.1153
$vec(\mathbf{A_3})$	0	0.0388	0.0455	0.0536		0	0.1564	0.2198	0.2630
	0	-0.0111	-0.0016	0.0083		0	0.0207	0.0286	0.0434
	0	-0.0021	0.0017	0.0061		0	0.1454	0.2404	0.3274
	0	-0.0034	-0.0014	0.0004		0	-0.1485	-0.0634	0.0384
	0	-0.0119	-0.0009	0.0119		0	-0.0453	0.0107	0.0459
	0	-0.0159	-0.0035	0.0081		0	0.0378	0.0787	0.1527
	0	-0.0074	-0.0002	0.0074	$\mathbf{\Psi}$	3	3.1668	3.2304	3.2954
	0	-0.0020	0.0026	0.0071		0	-0.0393	0.0112	0.1700
	0	-0.0098	0.0005	0.0122		2.5	1.8627	2.1368	2.4718
	0	-0.0100	0.0001	0.0117		0	-0.0699	0.0350	0.2396
	0	-0.0150	-0.0042	0.0082					
	0	-0.0070	0.0005	0.0088					
	0	-0.0097	-0.0004	0.0112					
	0	-0.0142	-0.0012	0.0106					
	0	-0.0100	0.0011	0.0121					
	0	-0.0098	0.0005	0.0130					
ν	15	14.7539	15.7235	16.3972					

Table 5.iii: Section 7.2: Distribution of the posterior mean of the parameters for $IWSV$ model, Real data, across $N = 100$ samples, Rolling window

Parameter	Prior mean	2.5% C.I.	Distribution mean	97.5% C.I.	Parameter	Prior mean	2.5% C.I.	Distribution mean	97.5% C.I.
$vech(\mathbf{C})$	0.3	0.3185	0.3531	0.3832	$vec(\mathbf{\Pi_1})$	0.2	0.0400	0.1544	0.2780
	0	-0.0371	-0.0066	0.0222		0	-0.0100	0.0065	0.0200
	0	-0.0132	0.0068	0.0336		0	0.0165	0.0310	0.0584
	0	-0.0169	-0.0063	0.0046		0	-0.0353	-0.0183	-0.0104
	0.3	0.3064	0.3458	0.3836		0	-0.1644	0.1977	0.4825
	0	-0.0209	-0.0008	0.0182		0.8	0.4511	0.5484	0.6119
	0	-0.0208	-0.0110	0.0013		0	-0.0454	0.0289	0.0486
	0.3	0.1913	0.2236	0.2495		0	-0.0092	0.0121	0.0332
	0	-0.0312	-0.0093	0.0070		0	-0.2187	-0.0084	0.1595
	0.3	0.0816	0.1003	0.1321		0	-0.0593	0.0253	0.1636
$vec(\mathbf{A_1})$	0.9	0.9910	0.9961	1.0023		0.8	1.0033	1.0719	1.1743
	0	-0.0179	-0.0046	0.0081		0	-0.0058	0.0160	0.0317
	0	-0.0027	0.0031	0.0101		0	-0.1244	0.0075	0.1431
	0	-0.0110	-0.0077	-0.0054		0	-0.3676	-0.1666	0.0647
	0	-0.0676	-0.0050	0.0482		0	-0.3851	-0.2110	-0.0573
	0.9	0.8813	0.9029	0.9404		0.8	1.0663	1.1334	1.2186
	0	-0.0137	0.0028	0.0188	$vec(\mathbf{\Pi_2})$	0	0.0633	0.1755	0.2832
	0	-0.0050	0.0027	0.0087		0	-0.0132	0.0032	0.0192
	0	-0.0348	0.0104	0.0534		0	0.0151	0.0240	0.0321
	0	-0.0157	0.0190	0.0502		0	-0.0240	-0.0170	-0.0113
	0.9	0.8688	0.8838	0.8970		0	-0.5463	-0.2745	0.0846
	0	-0.0285	-0.0216	-0.0121		0	0.0916	0.1512	0.2574
	0	-0.0358	-0.0048	0.0249		0	-0.0128	0.0363	0.1100
	0	-0.0418	-0.0103	0.0239		0	-0.0131	0.0066	0.0223
	0	-0.0314	-0.0085	0.0217		0	-0.5758	-0.3562	-0.1639
	0.9	0.7857	0.8007	0.8251		0	-0.1191	0.0599	0.1634
$vec(\mathbf{A_2})$	0	0.0331	0.0410	0.0511		0	-0.5216	-0.4447	-0.3696
	0	-0.0321	-0.0015	0.0264		0	-0.0522	-0.0314	-0.0133
	0	-0.0128	-0.0005	0.0094		0	0.0263	0.2298	0.4928
	0	-0.0042	-0.0015	0.0012		0	0.0015	0.2249	0.3491
	0	-0.0109	-0.0002	0.0082		0	-0.1277	-0.0247	0.0834
	0	-0.0150	-0.0030	0.0086		0	-0.4164	-0.2914	-0.1717
	0	-0.0107	-0.0012	0.0084	$vec(\mathbf{\Pi_3})$	0	-0.2182	-0.0745	0.0015
	0	-0.0036	0.0014	0.0050		0	0.0062	0.0231	0.0419
	0	-0.0099	0.0005	0.0097		0	-0.0056	0.0027	0.0101
	0	-0.0088	0.0004	0.0098		0	-0.0138	-0.0074	-0.0009
	0	-0.0144	-0.0029	0.0110		0	-0.3420	-0.1868	-0.0281
	0	-0.0086	-0.0011	0.0060		0	0.0505	0.1847	0.2479
	0	-0.0095	0.0000	0.0097		0	-0.0340	-0.0042	0.0250
	0	-0.0106	0.0004	0.0094		0	0.0027	0.0182	0.0298
	0	-0.0093	0.0003	0.0088		0	-0.1851	0.0293	0.2204
	0	-0.0079	0.0003	0.0094		0	-0.3163	-0.1617	0.0951
$vec(\mathbf{A_3})$	0	0.0342	0.0414	0.0486		0	0.1535	0.2405	0.3026
	0	-0.0263	0.0020	0.0289		0	0.0320	0.0602	0.0965
	0	-0.0119	0.0006	0.0126		0	-0.0743	0.0807	0.2953
	0	-0.0037	-0.0018	0.0003		0	-0.4598	-0.2316	0.1492
	0	-0.0127	-0.0008	0.0116		0	-0.0681	0.0543	0.1660
	0	-0.0135	-0.0021	0.0105		0	0.0391	0.0917	0.1640
	0	-0.0102	-0.0010	0.0126	$\mathbf{\Psi}$	3	2.9755	3.2153	3.4335
	0	-0.0028	0.0012	0.0056		0	-0.0780	0.0484	0.2550
	0	-0.0086	0.0000	0.0088		2.5	1.5758	1.8761	2.2094
	0	-0.0084	0.0008	0.0093		0	-0.1144	-0.0045	0.2037
	0	-0.0128	-0.0034	0.0041					
	0	-0.0060	0.0002	0.0056					
	0	-0.0117	-0.0014	0.0070					
	0	-0.0138	-0.0012	0.0103					
	0	-0.0110	0.0004	0.0104					
	0	-0.0080	0.0001	0.0088					
ν	15	14.8556	17.0125	19.2372					

Table 6.i: Section 7.2: Posterior distribution of the parameters for $Clark$ model, Real data, 1st sample, Recursive window

Parameter	Prior mean	2.5% C.I.	Posterior mean	97.5% C.I.	Parameter	Prior mean	2.5% C.I.	Posterior mean	97.5% C.I.
b_{21}	0	-0.0035	0.0340	0.0706	$vec(\mathbf{\Pi_1})$	0.2	0.0731	0.2563	0.4417
b_{31}	0	-0.0303	-0.0133	0.0013		0	-0.0431	0.0050	0.0535
b_{32}	0	0.0225	0.0537	0.0786		0	-0.0095	0.0040	0.0192
b_{41}	0	0.0298	0.0422	0.0543		0	-0.0445	-0.0275	-0.0113
b_{42}	0	0.0015	0.0285	0.0562		0	-0.1109	0.3320	0.7760
b_{43}	0	0.0221	0.0758	0.1242		0.8	0.3110	0.5186	0.7296
φ_1	doesn't exist	0.2569	0.5142	0.9586		0	-0.0461	-0.0149	0.0154
φ_2	doesn't exist	0.2716	0.5229	0.9416		0	-0.0440	-0.0043	0.0354
φ_3	doesn't exist	0.3263	0.6611	1.2517		0	-0.8804	-0.3001	0.2954
φ_4	doesn't exist	0.2419	0.4893	0.9293		0	-0.1741	0.1089	0.4020
						0.8	1.1264	1.3160	1.5004
						0	-0.0403	0.0288	0.0889
						0	-0.7122	0.0270	0.7621
						0	-0.8124	-0.2888	0.2411
						0	-0.3009	-0.1286	0.0573
						0.8	0.9702	1.1552	1.3412
					$vec(\mathbf{\Pi_2})$	0	-0.1059	0.0762	0.2561
						0	-0.0319	0.0255	0.0809
						0	-0.0074	0.0065	0.0220
						0	-0.0373	-0.0216	-0.0066
						0	-0.7804	-0.2908	0.1838
						0	-0.0849	0.1416	0.3713
						0	-0.0565	-0.0275	0.0051
						0	-0.0463	-0.0075	0.0305
						0	-0.8763	-0.1833	0.5165
						0	-0.4712	-0.0387	0.3979
						0	-0.8819	-0.6029	-0.3166
						0	-0.1286	-0.0536	0.0286
						0	-0.4021	0.3384	1.0790
						0	-0.4346	0.1610	0.7570
						0	-0.0309	0.2028	0.4361
						0	-0.7155	-0.4575	-0.1950
					$vec(\mathbf{\Pi_3})$	0	-0.3412	-0.1761	-0.0084
						0	-0.0338	0.0140	0.0640
						0	-0.0147	0.0006	0.0169
						0	-0.0186	-0.0057	0.0069
						0	-0.6458	-0.2553	0.1521
						0	-0.0510	0.1404	0.3184
						0	-0.0190	0.0039	0.0344
						0	-0.0108	0.0225	0.0550
						0	-0.4066	0.2528	0.9053
						0	-0.3194	0.0160	0.3600
						0	0.0828	0.2758	0.4662
						0	-0.0340	0.0250	0.0853
						0	-0.5465	0.1582	0.8685
						0	-0.4387	0.0452	0.5377
						0	-0.2472	-0.0772	0.0851
						0	0.0155	0.1813	0.3413
					$\mathbf{\Psi}$	3	2.9053	3.2772	3.6230
						0	-0.2167	0.1691	0.5383
						2.5	1.0637	2.2502	3.5255
						0	-0.0718	0.2747	0.6080

Table 6.ii: Section 7.2: Distribution of the posterior mean of the parameters for $Clark$ model, Real data, across $N = 100$ samples, Recursive window

Parameter	Prior mean	2.5% C.I.	Distribution mean	97.5% C.I.	Parameter	Prior mean	2.5% C.I.	Distribution mean	97.5% C.I.
b_{21}	0	0.0271	0.0343	0.0413	$vec(\mathbf{\Pi_1})$	0.2	0.2193	0.2416	0.2597
b_{31}	0	-0.0226	-0.0186	-0.0140		0	0.0055	0.0222	0.0398
b_{32}	0	0.0316	0.0400	0.0534		0	0.0032	0.0051	0.0073
b_{41}	0	0.0361	0.0403	0.0447		0	-0.0321	-0.0243	-0.0195
b_{42}	0	0.0185	0.0229	0.0317		0	0.0629	0.1623	0.3159
b_{43}	0	0.0807	0.0938	0.1072		0.8	0.5087	0.5298	0.5612
φ_1	doesn't exist	0.3160	0.4001	0.5112		0	-0.0180	-0.0082	-0.0026
φ_2	doesn't exist	0.3743	0.4450	0.5259		0	-0.0145	0.0080	0.0165
φ_3	doesn't exist	0.5106	0.5640	0.6417		0	-0.2665	-0.1119	-0.0251
φ_4	doesn't exist	0.3213	0.3985	0.5030		0	-0.0825	-0.0068	0.1260
						0.8	1.2945	1.3557	1.4073
						0	0.0015	0.0144	0.0371
						0	-0.0477	-0.0011	0.0499
						0	-0.3630	-0.3034	-0.2356
						0	-0.1271	-0.1025	-0.0671
						0.8	1.1025	1.1652	1.2153
					$vec(\mathbf{\Pi_2})$	0	0.0829	0.1489	0.1932
						0	0.0128	0.0194	0.0259
						0	0.0060	0.0081	0.0102
						0	-0.0233	-0.0178	-0.0138
						0	-0.2946	-0.1851	-0.1330
						0	0.1322	0.1765	0.2121
						0	-0.0275	-0.0133	-0.0036
						0	-0.0112	-0.0003	0.0070
						0	-0.3450	-0.1816	-0.0826
						0	-0.0759	0.0510	0.1085
						0	-0.6551	-0.5973	-0.5603
						0	-0.0594	-0.0419	-0.0344
						0	0.2069	0.3600	0.4790
						0	0.0497	0.2284	0.3607
						0	0.1230	0.1734	0.2045
						0	-0.4429	-0.3629	-0.2985
					$vec(\mathbf{\Pi_3})$	0	-0.1724	-0.1322	-0.0989
						0	0.0030	0.0204	0.0341
						0	-0.0018	0.0005	0.0029
						0	-0.0083	-0.0024	0.0007
						0	-0.2876	-0.2446	-0.1950
						0	0.1273	0.1985	0.2413
						0	0.0039	0.0079	0.0128
						0	0.0165	0.0231	0.0316
						0	0.0396	0.1384	0.2786
						0	-0.1424	-0.0560	0.0411
						0	0.1336	0.2178	0.2769
						0	0.0236	0.0364	0.0559
						0	0.0896	0.1705	0.2605
						0	-0.1737	-0.0393	0.1113
						0	-0.0852	-0.0694	-0.0389
						0	0.0537	0.1045	0.1890
					$\mathbf{\Psi}$	3	3.1523	3.2340	3.3483
						0	-0.1078	-0.0359	0.1655
						2.5	1.9251	2.2429	2.6220
						0	-0.0373	0.0878	0.3491

Table 6.iii: Section 7.2: Distribution of the posterior mean of the parameters for $Clark$ model, Real data, across $N = 100$ samples, Rolling window

Parameter	Prior mean	2.5% C.I.	Distribution mean	97.5% C.I.	Parameter	Prior mean	2.5% C.I.	Distribution mean	97.5% C.I.
b_{21}	0	0.0116	0.0332	0.0508	$vec(\mathbf{\Pi_1})$	0.2	0.1299	0.1831	0.2565
b_{31}	0	-0.0392	-0.0255	-0.0145		0	0.0010	0.0248	0.0539
b_{32}	0	-0.0440	-0.0130	0.0558		0	0.0062	0.0240	0.0516
b_{41}	0	0.0301	0.0360	0.0442		0	-0.0321	-0.0190	-0.0110
b_{42}	0	0.0053	0.0163	0.0315		0	-0.3065	0.0323	0.2289
b_{43}	0	0.0861	0.1064	0.1342		0.8	0.4393	0.5232	0.5880
$\varphi_{\mathbf{1}}$	doesn't exist	0.4483	0.4846	0.5617		0	-0.0245	0.0125	0.0449
$\varphi_{\mathbf{2}}$	doesn't exist	0.4900	0.5326	0.5793		0	-0.0093	0.0244	0.0372
$\varphi_{\mathbf{3}}$	doesn't exist	0.5823	0.6738	0.7938		0	-0.2616	-0.1561	-0.0367
$\varphi_{\mathbf{4}}$	doesn't exist	0.4382	0.4754	0.5246		0	-0.1113	-0.0291	0.1231
						0.8	1.0559	1.1462	1.2749
						0	-0.0004	0.0237	0.0350
						0	-0.0360	0.0300	0.1174
						0	-0.3795	-0.2548	-0.0874
						0	-0.3983	-0.2249	-0.0638
						0.8	1.0626	1.1299	1.2290
					$vec(\mathbf{\Pi_2})$	0	0.0678	0.1641	0.2412
						0	-0.0113	0.0101	0.0325
						0	0.0004	0.0157	0.0258
						0	-0.0215	-0.0162	-0.0129
						0	-0.4334	-0.2454	-0.0504
						0	0.1170	0.1688	0.2611
						0	-0.0253	0.0368	0.0770
						0	-0.0142	-0.0022	0.0172
						0	-0.3906	-0.2491	-0.1371
						0	-0.0668	0.0364	0.1110
						0	-0.6243	-0.5301	-0.4509
						0	-0.0531	-0.0336	-0.0188
						0	0.0121	0.2567	0.4570
						0	0.0297	0.2753	0.4241
						0	-0.1188	0.0901	0.2890
						0	-0.4588	-0.3234	-0.1831
					$vec(\mathbf{\Pi_3})$	0	-0.1717	-0.0732	-0.0288
						0	0.0055	0.0223	0.0481
						0	-0.0037	0.0067	0.0179
						0	-0.0119	-0.0070	-0.0031
						0	-0.2680	-0.1150	-0.0027
						0	0.1423	0.2313	0.2925
						0	-0.0007	0.0213	0.0496
						0	0.0005	0.0146	0.0272
						0	-0.0211	0.1319	0.2636
						0	-0.1609	-0.0717	0.0500
						0	0.1379	0.2394	0.3253
						0	0.0311	0.0562	0.0833
						0	0.0640	0.1231	0.1940
						0	-0.4318	-0.1993	0.1593
						0	-0.2045	-0.0389	0.1679
						0	0.0603	0.1250	0.1950
					$\mathbf{\Psi}$	3	2.9878	3.2183	3.5211
						0	-0.1435	0.0376	0.2415
						2.5	1.4283	1.7693	2.0976
						0	-0.1917	-0.0455	0.2607

Figure 23.i: IWSV, Posterior of parameters across $N = 100$ recursive sample windows

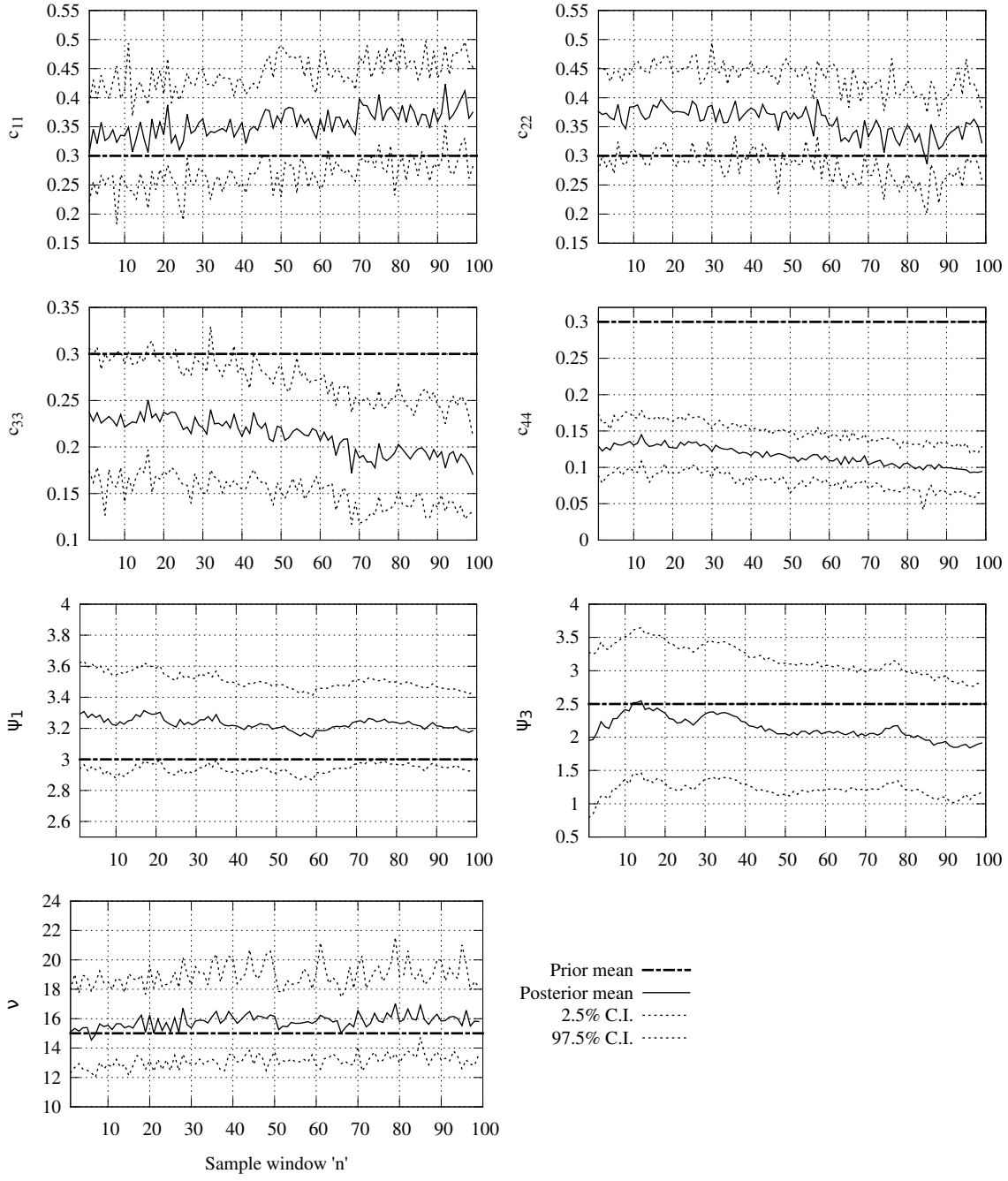


Figure 23.ii: IWSV, Posterior of parameters across $N = 100$ rolling sample windows

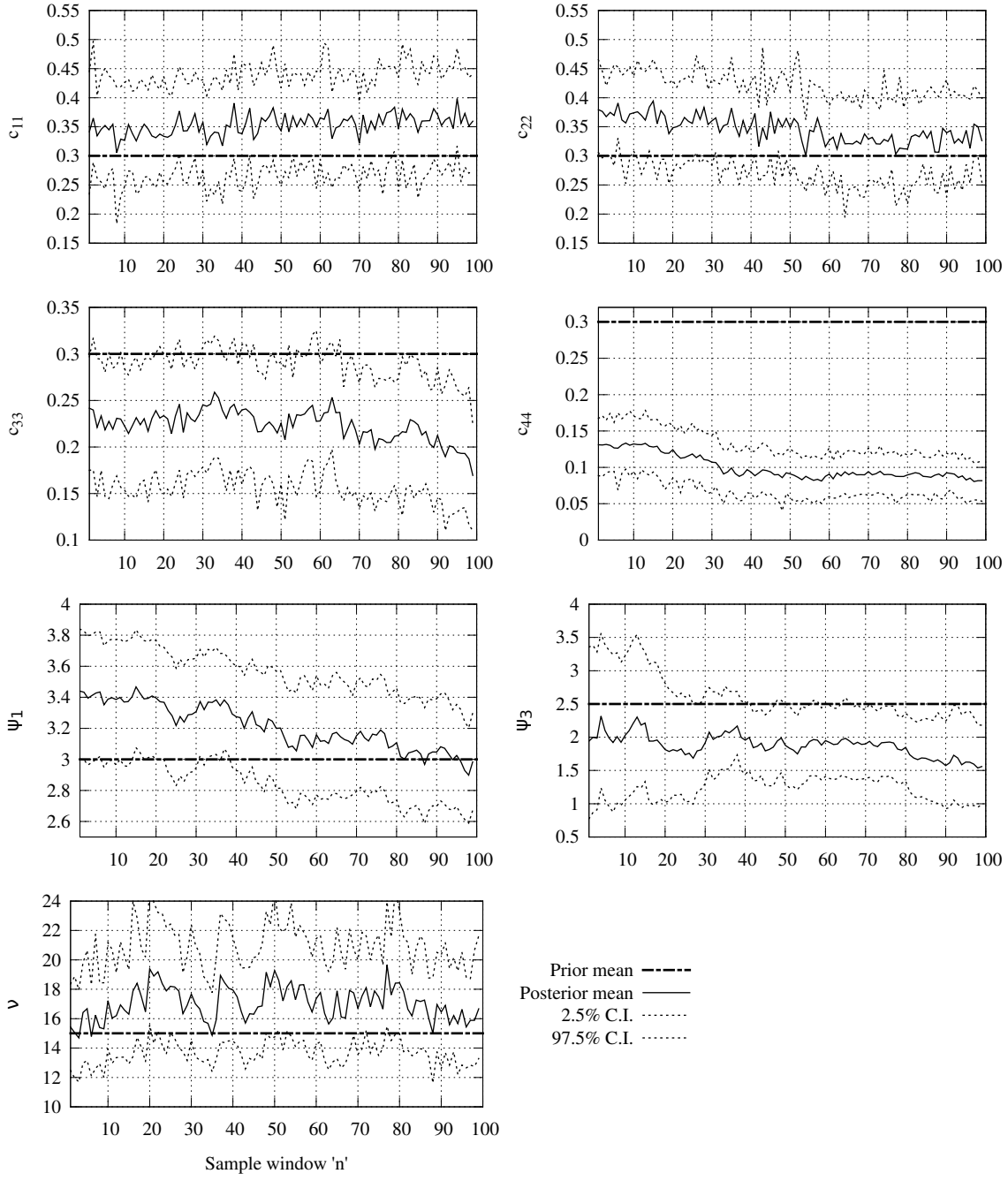


Figure 24.i: Clark, Posterior of parameters across $N = 100$ recursive sample windows

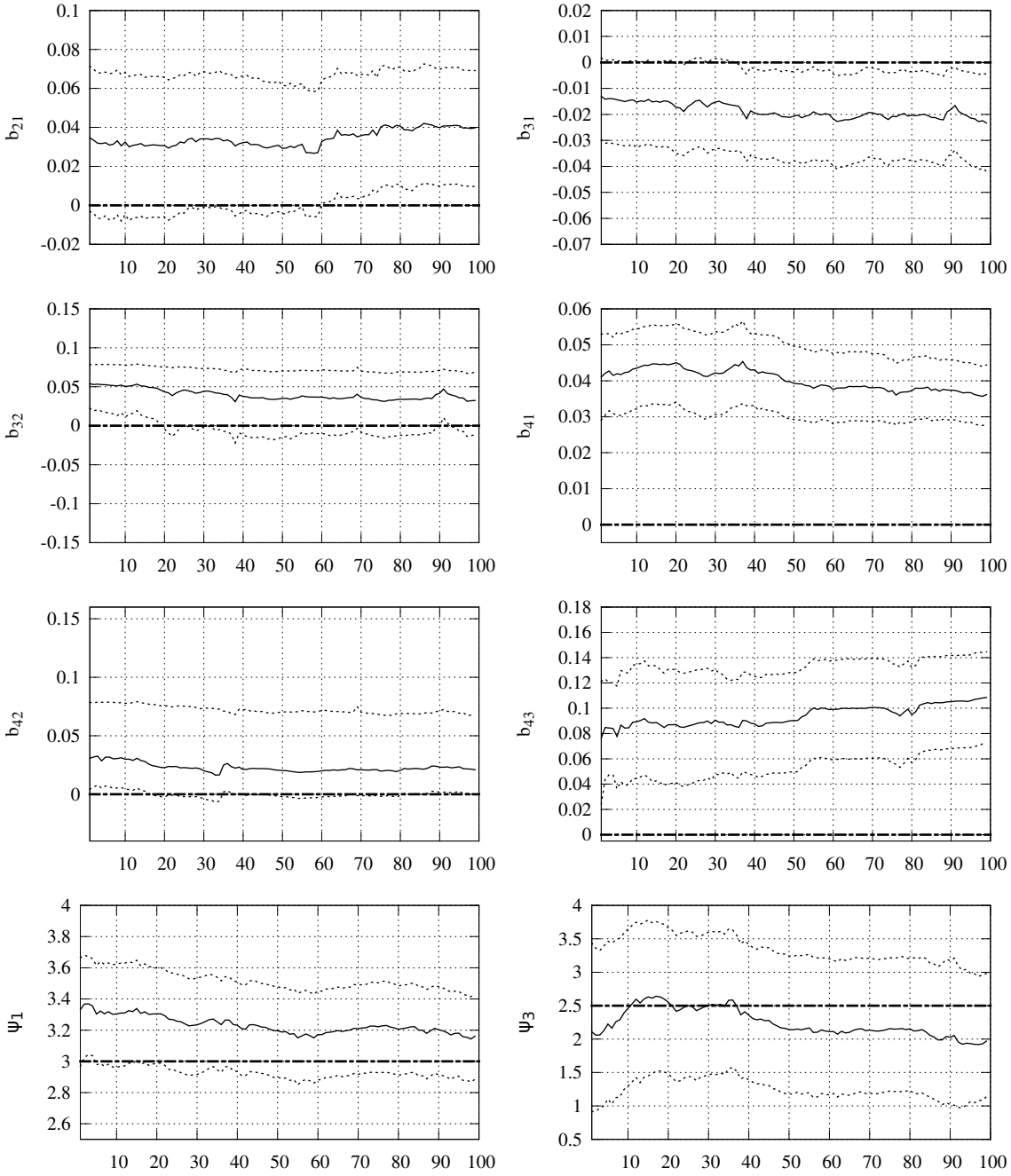


Figure 24.ii: Clark, Posterior of parameters across $N = 100$ rolling sample windows